Wireless Pers Commun
DOI 10.1007/s11277-017-4674-5

Data Aggregation in Wireless Sensor Networks: Previous Research, Current Status and Future Directions

Sukhchandan Randhawa[1] · Sushma Jain[1]

Abstract Wireless sensor networks (WSNs) consist of large number of small sized sensor nodes, whose main task is to sense the desired phenomena in a particular region of interest. These networks have large number of applications such as habitat monitoring, disaster management, security and military etc. Sensor nodes are very small in size and have limited processing capability as these nodes have very low battery power. WSNs are also prone to failure, due to low battery power constraint. Data aggregation is an energy efficient technique in WSNs. Due to high node density in sensor networks same data is sensed by many nodes, which results in redundancy. This redundancy can be eliminated by using data aggregation approach while routing packets from source nodes to base station. Researchers still face trouble to select an efficient and appropriate data aggregation technique from the existing literature of WSNs. This research work depicts a broad methodical literature analysis of data aggregation in the area of WSNs in specific. In this survey, standard methodical literature analysis technique is used based on a complete collection of 123 research papers out of large collection of 932 research papers published in 20 foremost workshops, symposiums, conferences and 17 prominent journals. The current status of data aggregation in WSNs is distributed into various categories. Methodical analysis of data aggregation in WSNs is presented which includes techniques, tools, methodology and challenges in data aggregation. The literature covered fifteen types of data aggregation techniques in WSNs. Detailed analysis of this research work will help researchers to find the important characteristics of data aggregation techniques and will also help to select the most suitable technique for data aggregation. Research issues and future research directions have also been suggested in this research literature.

✉ Sukhchandan Randhawa
sukhchandan@thapar.edu

Sushma Jain
sjain@thapar.edu

[1] Computer Science and Engineering Department, Thapar University, Patiala, Punjab 147004, India

Published online: 28 July 2017

Keywords Data aggregation · Wireless sensor networks · Energy efficiency · Quality of service

1 Introduction

Wireless sensor networks (WSNs) consist of a large number of sensor nodes deployed in region of interest for specific applications. WSNs are different kind of networks having their importance in areas such as environment monitoring, military applications, health care applications, industrial process control, home intelligence, security and surveillance etc. These nodes are generally of small size and have computation, communication and sensing capabilities. Sensor nodes communicate via short range radio signals and collaborate among themselves to accomplish the common tasks [1, 2]. However, sensor nodes have limited bandwidth, power, memory, processing resources and limited lifetime. The main task of sensor node is to sense the target phenomena such as heat, light, temperature and then report that data to the host controller or sink in the form of query response. In WSNs, energy consumption is less for computation as compared to data transmission [3]. Rather than sending the sensed data each time to the sink node individually, if data is first collected and aggregated using aggregate functions such as sum(), avg() etc. and then forwarded to the sink, a lot of energy will be saved. In WSNs, data aggregation is a process of collecting and combining the useful information in a particular region of interest. The effectiveness of the communication among nodes depends on the data aggregation technique being used. Data aggregation can be considered as a fundamental processing procedure to reduce energy consumption and to save the limited resources. An effective data aggregation technique can enhance energy efficiency and network lifetime [4]. In this paper, a survey on various techniques for data aggregation is presented. Data aggregation can also be useful when multiple nodes sense the same phenomenon due to high node density (also known as *data overlapping problem*) [5]. The working of data aggregation in WSN is shown in the Fig. 1.

Data aggregation technique can be used to aggregate the sensor readings as shown in Fig. 1 [6]. First of all, data is collected from various nodes. The various algorithms like low energy adaptive clustering hierarchy (LEACH), centralized approach, Tiny AGgregation (TAG) etc. are used to aggregate the sensor data coming from the sensor nodes. The input for an aggregation algorithm is the sensor readings taken from various nodes and output is the aggregated data. Further an efficient path is selected to transfer the aggregated data to the sink node [7, 8].

A routing protocol is used to decide the best route suitable for sending data to the sink from a source node [2, 9]. There is a high probability of failure in WSNs due to energy constraints, link failures and coverage hole problem etc. Due to dynamic environment in

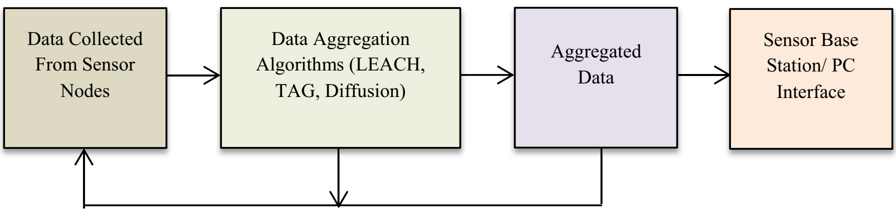


Fig. 1 Data aggregation process

which a WSN is deployed, it is not sufficient to rely on a single path between the source and destination. In order to deal with such problems, multiple paths between source and destination need to be discovered, so that if one path fails, data can be transmitted to another discovered path. So, there is a need to develop a routing protocol which is efficient in terms of resource consumption (energy resource is the major concern) and perform data aggregation by discovering multiple paths between source and destination. The security issues, energy consumption, data confidentiality, latency, integrity and data aggregation become vital when the sensor network is deployed in a hostile environment [10].

1.1 Motivation for Research

- Data aggregation is a process of combining the sensor data using aggregation approaches. Consequently, this study emphasizes on data aggregation approaches based on different criteria.
- The necessity of methodical literature survey has been recognized after considering progressive research in data aggregation in WSNs. Therefore, based on broad and methodical search in existing database, the available research is summarized and research challenges for future research are also presented.

1.2 Related Surveys

Rajagopalan and Varshney [11] has done an innovative literature review in the field of data aggregation in WSNs. Jesus et al. [12] reviews the state of the art on distributed data aggregation algorithms and characterize the different types of aggregation functions. Still the research is persistently growing in the field of data aggregation. There is a necessity of methodical literature survey to evaluate and integrate the existing research presented in this field. This research presents a methodical literature survey to evaluate and discover the research challenges based on available existing research in the field of data aggregation in WSNs.

1.3 Paper Organization

The organization of rest of this paper is as follows: Data aggregation taxonomy is presented in Sect. 2. Section 3 presents the background of data aggregation along with evolution and types of queries used in data aggregation. Section 4 describes the review technique used to find and analyze the available existing research, research questions and searching criteria. Section 5 describes the categorization of data aggregation techniques in WSNs and their comparisons and taxonomy. Section 6 describes the perspective solution for data aggregation. Section 7 presents tools and platforms used for data aggregation in WSNs and a discussion is presented in Sect. 8. Section 9 presents the extraction outcomes of the methodical literature survey. Future research directions and conclusions are described in Sect. 10. Note: A glossary of acronyms used in this paper can be found in Appendix 1.

2 Data Aggregation Taxonomy

The data aggregation taxonomy like network lifetime, energy efficiency, data accuracy, latency and data aggregation rate [6, 13–22] are used in this survey are described below:

(a) *Energy Efficiency* Every sensor should consume same amount of energy in every data gathering round in ideal situation but in real situations, sensor nodes consume different amount of energy for data transmission. Data aggregation technique is energy efficient if it provides the maximum functionality with minimum energy consumption in WSNs. Energy efficiency is a ratio of amount of data successfully transferred in a sensor network to total energy consumed to transfer those data. Equation 1 is used to calculate energy efficiency.

$$Energy\ Efficiency_i = \sum_{i=1}^{n} \left(\frac{Amount\ of\ data\ successfully\ transferred\ in\ a\ sensor\ network}{Total\ energy\ consumed\ to\ transfer\ those\ data} \right) \tag{1}$$

where n is the number of sensors nodes in a sensor network.

(b) *Network Lifetime* Network lifetime can be defined as number of data aggregation rounds finished till first sensor node is exhausted of its energy. In other words, it is defined as the time until the first sensor node or group of sensor nodes in the network runs out of energy (battery power) or the time (number of rounds) of network disconnection due to the failure of one or more sensor as (Eq. 2):

$$NL_n^n = \min_{v \in V} NL_v \tag{2}$$

where the network lifetime NL_n^n ends as soon as the first node fails, with NL_v being the lifetime of node v and V is the node set excluding the sink node.

c) *Data Accuracy* It is defined in different contexts, based on the application for which sensor network is designed. For example, the close estimate of target location at the sink defines the data accuracy in target localization problem. Data accuracy is defined as the ratio of amount of data transferred successfully to Total amount of data sent as (Eq. 3):

$$Data\ Persistence = \frac{Amount\ of\ data\ transferred\ successfully}{Total\ amount\ of\ data\ sent} \tag{3}$$

d) *Latency* It is the amount of time between the data produced at the source nodes and the data packets received at the sink. In other words, latency can be a defined as a time difference between sending the data and receiving the data by a sensor node. Equation 4 is used to calculate Latency as follows:

$$Latency_i = \sum_{i=1}^{n} (Time\ of\ receiving\ data - Time\ of\ sending\ data) \tag{4}$$

where n is the number of sensors nodes in a sensor network.

e) *Data Aggregation Rate* In WSNs, data aggregation is a process of collecting and combining the useful information in a particular region of interest. Data aggregation can be considered as a fundamental processing procedure to reduce energy consumption and to save the limited resources and is defined in terms of data aggregation rate. Data aggregation rate is defined as the ratio of amount of data aggregated successfully to total amount of data sensed as (Eq. 5).

$$Data\,Aggregation\,Rate = \frac{Amount\,of\,data\,aggregated\,successfully}{Total\,amount\,of\,data\,sensed} \times 100 \qquad (5)$$

3 Background: Previous Research

In the beginning, the aspects leading to data aggregation are discussed. After this, the data aggregation algorithms are summarized and categorization of these algorithms is done. Then several facts are discovered as to know why data aggregation is beneficial frequently [23, 24]. The working of WSN architecture model is illustrated in Fig. 2.

First of all, nodes are selected and distributed among different clusters. Parameters like energy consumption, memory and bandwidth etc. are considered to calculate the number of nodes that will participate in a cluster. After this, one Cluster Head (CH) is nominated from each cluster [25, 26]. The responsibility of CH includes supervision of all nodes inside cluster and gathering the data from every node in a cluster and transfers this data to adjacent CH for updation [6]. Cluster will be allotted to newly arrived nodes. Further collection of data and various user queries are checked by data aggregation technique and then query processor transforms these queries into low level format [27]. Database server is used to store the gathered and aggregated data. Lastly data cube technique is used to aggregate all the data and send aggregated data to base station for further use.

3.1 Data Aggregation Evolution

The evolution of data aggregation techniques (DATs) is found from year 2002 to 2015 in WSNs. Further remarkable focus of study (FoS) of data aggregation by evolution of WSNs across the various years are described in data aggregation evolution as shown in Fig. 3. In year 2002, network lifetime [13] and network density [28] data aggregation techniques have been presented which focused on network lifetime and energy.

In year 2003, dynamic data aggregation [29] technique is proposed which focused on network lifetime. In year 2004, adaptive data aggregation technique [30] has been presented which focused on energy consumption and transmission delay. In year 2005,

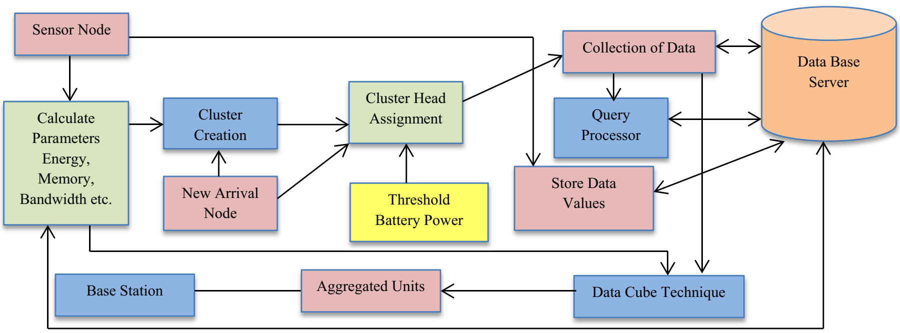


Fig. 2 Data aggregation and data gathering architecture in WSN

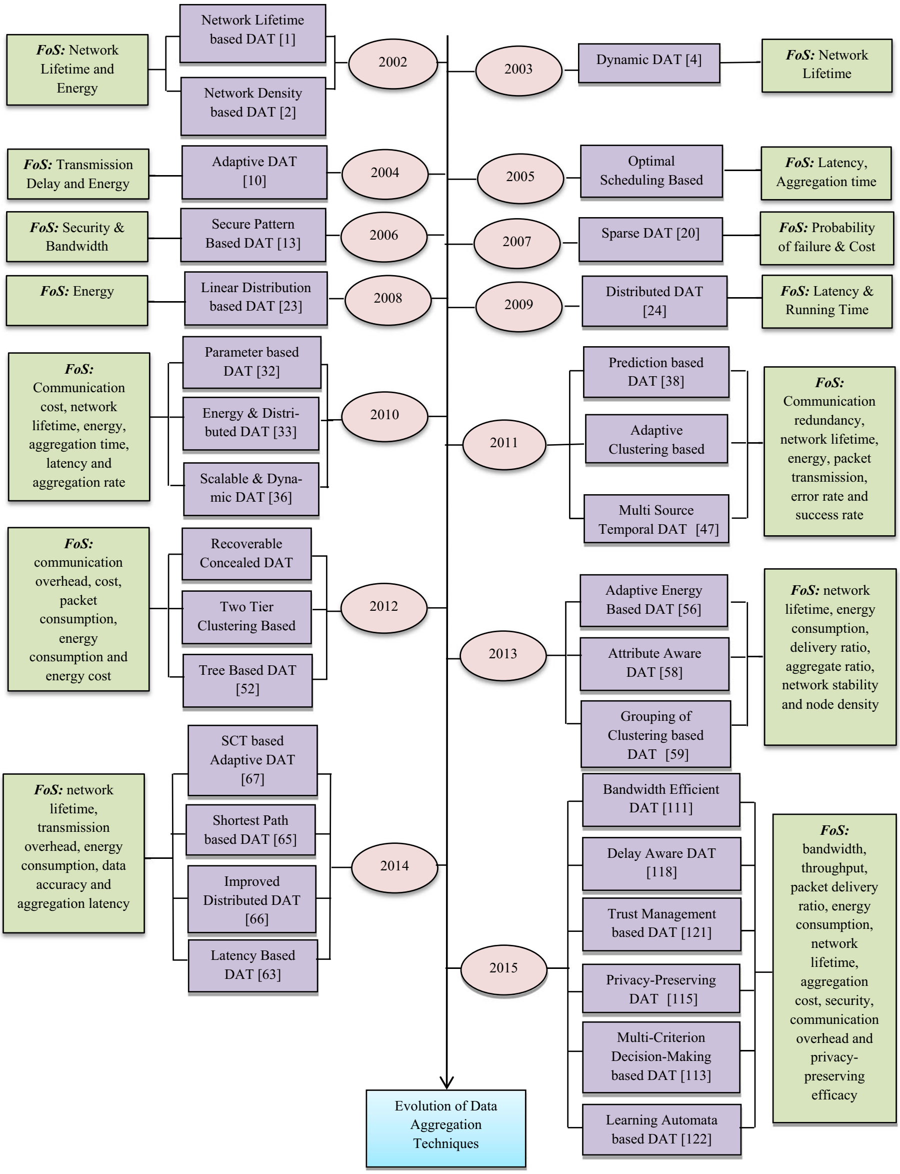


Fig. 3 Evolution of data aggregation techniques

optimal scheduling based data aggregation technique [31] has been presented which focused on aggregation time and latency. In 2006, secure pattern based data aggregation technique [32] has been presented which focused on security and bandwidth. In year 2007, sparse data aggregation technique [33] has been presented which focused on cost and probability of failure.

In year 2008, linear distribution based data aggregation technique [10] has been presented which focused on energy. In year 2009, distributed data aggregation technique [34] has been presented which focused on latency and running time. In year 2010, parameters based [35], energy oriented distributed [36] and scalable and dynamic [37] data aggregation techniques have been presented which focused on communication cost, network lifetime, energy, aggregation time, latency and aggregation rate. In year 2011, prediction [38], adaptive clustering [39] and multi-source temporal [40] based data aggregation techniques have been presented which focused on communication redundancy, network lifetime, energy, packet transmission, error rate and success rate. In year 2012, recoverable concealed [41], two tier clustering based [42] and tree based [43] data aggregation techniques have been presented which focused on communication overhead, cost, packet consumption, energy consumption and energy cost. In year 2013, adaptive energy based [44], attribute aware [45] and grouping of clustering based [46] data aggregation techniques have been presented which focused on network lifetime, energy consumption, delivery ratio, aggregate ratio, network stability and node density. In year 2014, shortest path based [19], semantic correlation tree (SCT) based adaptive [47], improved distributed [48] and latency based [18] data aggregation techniques have been presented which focused on network lifetime, transmission overhead, energy consumption, data accuracy and aggregation latency.

Recently in year 2015, bandwidth efficient [49], delay aware [50], trust management [51], multi-criterion decision-making [52] and learning automata based [53] data aggregation techniques have been proposed which focused on bandwidth, throughput, packet delivery ratio, energy consumption, network lifetime, aggregation cost, security, communication overhead and privacy-preserving efficacy.

3.2 Queries in Data Aggregation

In data aggregation, a query is generated at base station or cluster head and is sent towards source nodes to sense the data required for query to answer and to revert back with a reply. There are generally three types of queries, which are generated while data aggregation in WSNs namely *simple query, complex query* and *event based query*.

(1) *Simple query* In these queries, a predicate clause for filtering the sensed data is used. In these queries, aggregate functions or aggregators such as avg(),sum(),min(),max() etc. are not used. Example: SELECT humidity FROM sensors WHERE node = 1;

(2) *Complex query* In this type of query, aggregate functions and subqueries are used. Example: SELECT humidity FROM sensor WHERE loc = (SELECT roomno WHERE floorno = 7);

(3) *Event Driven Query* In this type of query, data is returned or reported from source nodes at periodic time intervals.

4 Review Technique

The methodical survey technique described in this research article has been taken from [54]. The stages of this literature review include creation of review framework, executing the survey, investigating the results of review, recording the review results and exploration of research challenges. Table 1 describes the list of research questions required to plan the survey in data aggregation in WSN. Table 2 describes the 932 research papers retrieved in manual search and electronic database search. Figure 4 describes the review technique

Table 1 Research questions and motivation

Review questions	Motivation
1. What is the current status of data aggregation? 2. How to cut down the communication and transmission cost? 3. How to maximize the network lifetime and improve resource utilization? 4. How to optimize the resource utilization and minimize the communication cost simultaneously? 5. How to deliver the data to destination in a timely fashion? 6. How to reduce energy consumption and its impact on network lifetime? 7. What other optimization techniques should be considered for efficient data aggregation? 8. What new rules should be required for effective data aggregation? 9. How to design the data aggregation algorithms to provide dynamic scalability and to reduce energy consumption? 10. How to develop a data aggregation technique based on QoS requirements? 11. What simulation tools are used for data aggregation and what parameters are considered? 12. How to validate the data aggregation technique through tools?	It aids in recognizing the data aggregation algorithms. Various data aggregation algorithms used in WSNs are reported. Various QoS parameters for data aggregation considered so far are stated according to their level of importance. The research challenges in terms of research questions discover the existing research which has been assessed. This study compares the different types of data aggregation techniques. For every type and subtype of data aggregation technique, various types of existing research have been presented. It will support in planning enhanced and extremely accessible approaches. The main aim of this review is to make data aggregation database for future research through standardization and benchmarking of relative investigation of existing research. It is mandatory to find out the number of research papers in each type data aggregation technique which helps to find the key research areas in subtypes of data aggregation techniques. Different research questions are used to identify the key research areas for future investigation in the field of data aggregation

used in this systematic review. Details of review technique used in this research work can be found in review paper [22].

4.1 Sources of Information

The following electronic database sources are searched for the presented review as recommended by [54]:

Springer (www.springerlink.com)	ACM Digital Library (www.acm.org/dl)
ScienceDirect (www.sciencedirect.com)	Wiley Interscience (www.Interscience.wiley.com)
Google Scholar (www.scholar.google.co.in)	HPC (www.hpcsage.com)
IEEE eXplore (www.ieeexplore.ieee.org)	Taylor & Francis Online (www.tandfonline.com)

4.2 Search Criteria

The keyword "data aggregation" is involved in the abstract of each research paper in every search. It is time consuming and general method for review. The various search strings used in this review are described in Table 2. This methodical literature survey included both types of research articles: quantitative as well as qualitative (from year 2002 to 2015). The basic research in this area is commenced in 2000 but rigorous development took place

Table 2 Search string

Sr. no.	Keywords	Synonyms	Dates	Content type
1	Data aggregation	Data aggregation technique	2002–2015	Journal, Conference, Workshop, Magazine, White paper and Transactions
2	WSN	Wireless sensor network		
3	Aggregation	Data aggregation in WSN		
4	Tools	Simulation tools in data aggregation		
5	Evolution	Review of existing research in data aggregation		
8	Analysis	Analysis of research gaps in data aggregation		
9	Comparison	Comparison of existing research		
10	QoS	Quality of service		
11	QoS and DA	Quality factors in data aggregation		
12	Energy, cost, network time	Data aggregation criteria in WSN		

after 2007. The research papers are included from journals, conferences, symposiums, workshops and white papers from industry along with technical reports. Exclusion criteria used at different stages is described in Fig. 4. An individual search mechanism is applied on some journals of Springer, Wiley, Taylor and Francis, Science Direct etc. Our search retrieved over 932 research articles as shown in Fig. 4, which were reduced to 781 research articles based on their titles, 322 research articles based on their abstracts and conclusions and 191 research articles based on full text. Then, these 191 research articles were investigated completely to find a final collection of 123 research articles through references investigation and eliminating common challenges based on the criterion of inclusion and exclusion.

5 Data Aggregation Techniques in WSNs: Current Status

This section discusses the current status of data aggregation in WSNs along with data aggregation methodology.

5.1 Data Aggregation Methodology

There are various methodologies according to which data aggregation is performed namely *centralized, tree based approach, cluster based approach* and *in-network aggregation* as shown in Fig. 5.

5.1.1 Centralized Approach

In this approach, each and every node transmits its sensed data to the central node, which is usually the most powerful node (in terms of resources such as energy, bandwidth etc.).

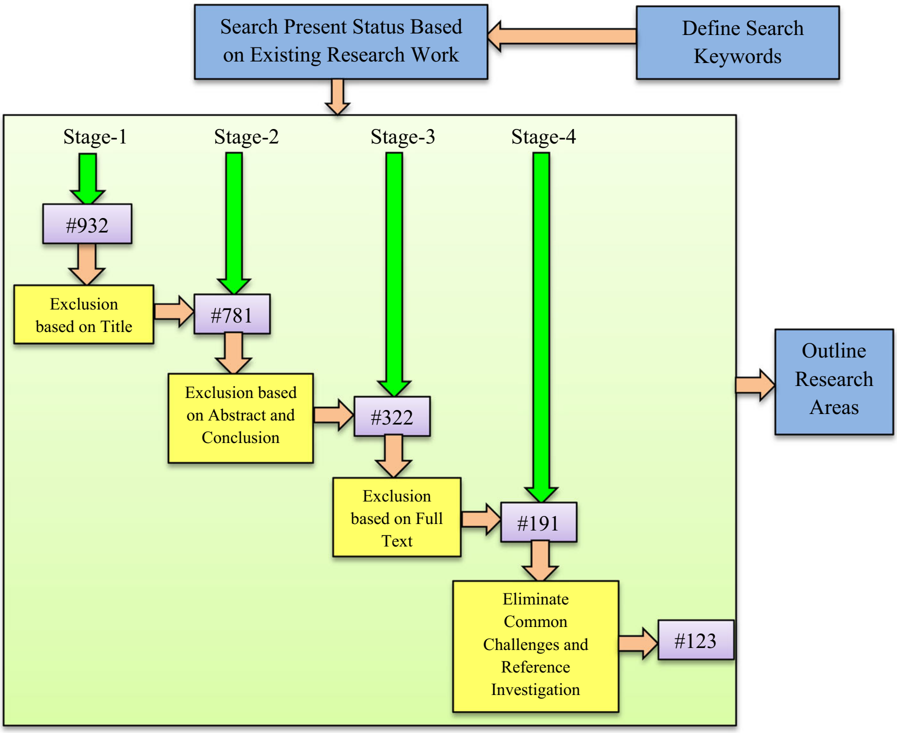


Fig. 4 Review technique used in this systematic review

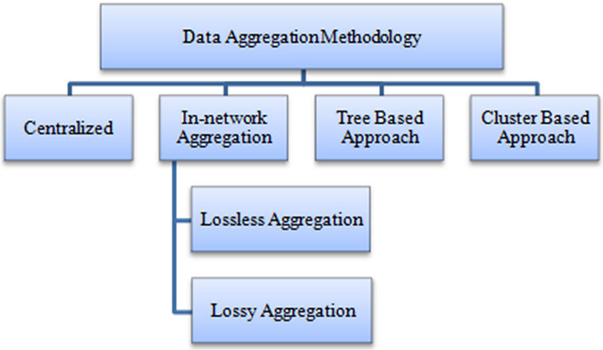


Fig. 5 Data aggregation methodologies

Address centric routing is used along with multi hop algorithm by considering cost metric at every intermediate node. The central node's task is to aggregate the data received from other nodes and report the same data to the base station. This approach suffers from the problem of high traffic due to large number of messages being transmitted.

5.1.2 In-Network Aggregation

In this methodology, data is processed at intermediate nodes in order to reduce consumption of critical resources such as energy, computation time etc. This approach also increases the network lifetime as it tries to reduce the energy consumption at every node. There are further two approaches for in-network aggregation: *lossy aggregation* with packet size reduction and *lossless data aggregation* without packet size reduction. In lossy aggregation, data is gathered from various source nodes and then some group function is applied over the gathered data such as sum(), count(), maximum() and minimum() etc. In this approach, size of the packet is reduced, as only calculated value of aggregate function is inserted into the packet after compression, rather than sending the whole packet of every node. Example, in forest fire monitoring system, average or maximum temperature reading is required in a timely manner. In such applications, lossy aggregation is required, as it responds in a timely manner to the base station. In lossless aggregation, every packet is merged into single packet without compression.

5.1.3 Tree Based Approach

In this approach, first an aggregation tree is built which is generally a minimum spanning tree. In this tree, root node acts as base station, leaf nodes act as source node and intermediate nodes act as parent nodes. Leaf node send their sensed node to their parent node in a path discovered between leaf node and base station as shown in Fig. 6.

5.1.4 Cluster-Based Approach

In this approach, region of interest is divided into number of clusters. Within each cluster, a cluster head is elected, whose main task is to aggregate the data. Each node senses the desired phenomena and reports the data to the cluster head of the same cluster rather than sending the data directly to the base station. Hence it saves a lot of energy in a network.

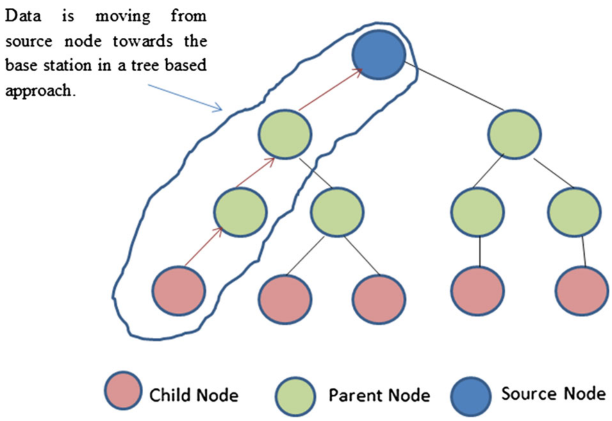


Fig. 6 Tree based approach for data aggregation

5.2 Data Aggregation Techniques

Data aggregation technique (DAT) plays an important role in gathering and aggregating useful information in WSNs [3].The various types of data aggregation algorithms are shown in Fig. 7. Table 3 presents a comparison of these techniques based on their common features.

5.2.1 Adaptive Data Aggregation Techniques

Adaptive data aggregation research work has been done by following authors. Lu et al. [14] proposed latency based energy efficient adaptive data aggregation technique i.e. DMAC. This technique which has three advantages over SMAC which are improved packet delivery latency, no more fixed duty cycle and less collision due to fixed synchronous duty cycle. Before this, all other MAC protocols suffer from the problem of *Data Forwarding Interruption problem*, which arises due to the nodes residing in sleep mode to which data is forwarded next. But this problem is removed in proposed DMAC, as it follows staggered active/sleep schedule, so that data can be forwarded in continuous flow in a multi-hop fashion. It uses data prediction mechanism along with *More-To-Send* packet, by which a parent node in a tree structure will get to know that whether its child node has more data to send or not. This approach ensures reduced energy consumption and improved latency, but DMAC may not be suitable in a burst traffic scenario.

He et al. [30] presented adaptive application independent data aggregation (AIDA) technique, in which all decisions of aggregation are taken by an intermediate layer between data link layer (MAC) and network layer independent of the application. So there is no need to rewrite the logic of aggregation specific to the application as it adjusts itself according to the varying network traffic conditions in a timely manner. There are two units in this proposed aggregation layer. First one is an *aggregation function unit* which aggregates the packets residing in an aggregation pool and selects on of the four packet formats namely broadcast, anycast, multicast and unicast. Depending upon the

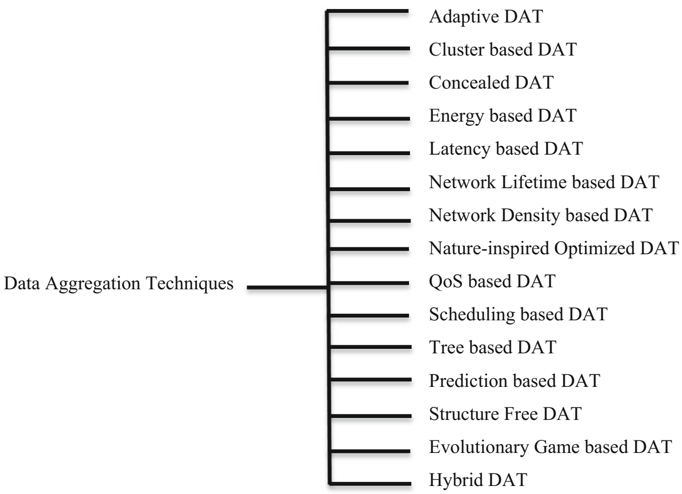


Fig. 7 Data aggregation techniques

Table 3 Traits of data aggregation techniques

Trait	Description
Sub type	Data aggregation techniques have been further divided into various subtypes
Routing protocol	Literature reported that different types of routing protocols are used in different Data Aggregation Techniques. Routing protocols identified from existing literature are distributed, chain based, tree based, hierarchical and cluster based
Objective function	An objective function of every data aggregation technique is specifically designed for a specific purpose of the mechanism. For example: To reduce latency and energy consumption. Mostly, the main objective function of the mechanism is the efficient aggregation of data by optimizing performance parameters
Optimizing parameter	Every data aggregation technique has their performance parameters specifically designed for a specific purpose of the result comparison. For example: Energy, cost, time, network lifetime, data accuracy etc
Merits	The advantages of data aggregation technique are described in this section
Operational environment	An operational environment is which data aggregation technique can be implemented and executed. Different type of tools used in existing data aggregation techniques for evaluation (described in Sect. 7) are discussed
Demerits/open issues	The disadvantages or any open issues of data aggregation technique are described in this section

requirement, the second unit i.e. *aggregation control unit* adapts the system according to the dynamic feedback and also controls degree of aggregation. It improves end to end delay and transmission energy but incurs feedback overhead. Chen et al. [55] proposed cluster based adaptive data aggregation (ADA) technique, in which the main focus is to shift the load of sensor nodes and cluster heads to the resource enriched sink side and to enhance reliability factor of network. Sink node determines the reporting frequency of nodes and aggregation frequency of cluster heads. The reliability of event features is represented by two reliability parameters namely spatial reliability and temporal reliability. Spatial reliability is determined by aggregation ratio whereas temporal reliability is determined by reporting frequency. It achieves the desired reliability, but a lot of effort is required in terms of two dimensional parameters of spatial–temporal correlation.

Virmani et al. [44] presented adaptive energy aware data aggregation tree (AEDT), in which a node with maximum energy is elected as an aggregator node. All other nodes except the sink node and aggregator node go to the sleep mode, so it reduces energy consumption. There is a threshold value for traffic, beyond which packets are selected according to the parent capacity. If the estimated traffic load is greater than the communication capacity of parent node then an OVERLOAD message is transmitted to the network. A memory table is also maintained in which all the discovered paths are stored. After t seconds (where t is real time delay), proposed tree is refreshed and new root node (parent) is elected if required. Shortest path and energy are the two important factors which are considered while routing. But overhead of maintaining a memory table of such a dynamic network is an overhead. Kumar and Rajkumar [47] proposed approximation based semantic correlation tree (SCT) data aggregation technique in which whole sensor field is divided into ring like structures. Each ring is further partitioned in terms of sectors. A sector head is selected in each sector, which works as an aggregation node whose task is to report the aggregated data from its respective sector to the base station. In this doorway algorithm, the base station is assumed to be in the center of the sensor field. It is based on *local estimation* model, in which it can generate a new reading based on the past history to reduce the network congestion and communication cost. The correctness of estimation model is verified by checking the estimation error. If the estimation error is not within the bounds, then this is called *anomaly*. As WSNs are randomly deployed, so finding optimal number of sectors, rings and number of nodes within each sector is not an easy task.

Rout and Ghosh [25] presented energy efficient adaptive data aggregation using network coding (ADANC). It is energy oriented cluster based data aggregation technique in which sensor nodes of a cluster are divided into two categories: *simple relay nodes* and *network coder nodes*. Network coder nodes act as *aggregator nodes*. Data aggregation is based on level of data correlation. If the level of data correlation factor is less in the received packets, then network encoding is performed. Otherwise traditional data aggregation is performed when the data correlation factor is high. Energy consumption is also reduced at node level with duty cycle. This technique is used to improve energy consumption by reducing communication cost and network lifetime but data latency is not considered.

Wu et al. [56] proposed robust dynamic data aggregation technique. It is a hybrid approach which consider varying network dynamics (as in dynamic aggregation) and robustness (which deals with packet loss). To achieve robustness virtual rings are used which transfers a packet to multiple nodes. For dynamic data aggregation, cross layer approach i.e. DA-MAC is used which decides when and where to aggregate the data. It is specifically design for event triggered based scenario, in which events can occur at unknown time and unknown regions in the covered area. The duty cycle of each node is asynchronous which means that each node has its own sleep/wake schedule. *Clear Channel Assessment* signal is used to check the medium status: busy or free. This technique is used to improve robustness, overall cost, duty cycle and delay but network lifetime can be improved.

Adaptive Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 8. In *energy and latency* based, data is aggregated with lesser energy consumption and latency. In approximation based approach, in order to reduce communication cost, correlation factor in the received data is considered. In *reliability* based data aggregation, the main focus is to shift the load of sensor nodes and cluster heads to the resource enriched sink side and to enhance reliability. In *energy based data aggregation*, there is a threshold value for each parent node in a tree, beyond which network is intimated by OVERLOAD message. In *application* independent data aggregation, all decisions of aggregation are taken by an intermediate layer between data link layer (MAC) and network layer independent of the application. In *energy and data correlation* based data aggregation, fuzzy logic is used to select the CH and distribute the equal amount of load among clusters to transfer data in optimal way with minimum energy consumption. *Robust dynamic* data aggregation is a hybrid approach which considers varying network dynamics (as in dynamic aggregation) and robustness (which deals with packet loss) using cross layer design.

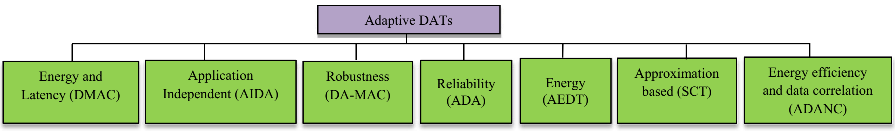


Fig. 8 Adaptive DATs taxonomy

5.2.2 Cluster Based Data Aggregation Techniques

Data aggregation research work based on cluster has been done by following authors. Chen et al. [55] proposed cluster based adaptive data aggregation (ADA) technique, which has already been discussed in Sect. 5.2.1 i.e. adaptive based data aggregation techniques.

Zheng et al. [57] presented distributed data aggregation mechanism. Its main objective is to solve clustered Slepian–Wolf coding (CSWC) problem to choose a disjoint potential cluster to cover the network and to increase the compression gain. It aims to maximize the global compression gain. Authors designed distributed optimal compression (DOC) clustering protocol and a low-complexity joint-coding technique to decrease the data redundancy generated by spatial correlation among various clusters. The proposed algorithm can find an optimal rate allocation within each cluster in order to minimize the intra-cluster communication cost. To perform Slepian–Wolf coding within a single cluster, an intra-cluster coding protocol has been presented. This technique also reduces communication cost and compression ratio but energy consumption is not considered. Jung et al. [39] proposed hybrid based data aggregation technique named *combined clustering based* data aggregation to enable dynamic aggregation by applying various techniques of clustering simultaneously. An adequate clustering technique is selected based on status of the network. In this proposed technique, first of all tree topology is created between the sink node and other nodes of a network in initialization phase. After this phase, next step is to apply cluster head static election algorithm. This is a hybrid approach in the sense, that it provides both static as well as dynamic clustering depending upon the network status. This technique improves network lifetime and energy consumption without considering latency as an optimization criterion.

Maraiya et al. [58] presented efficient cluster head selection scheme for data aggregation (ECHSSDA) which uses the concept of selection of cluster head and formation of cluster. This technique reduces overhead of clustering by electing a cluster head along with an *Associate cluster head*. In clustering, cluster head can die soon as it is overloaded with tasks such as sending, receiving and computations. So when a CH dies, a new election is performed to elect new CH and also reclustering is performed to avoid *hot spot problem*. But, this overhead can be avoided by using the concept of an associate cluster head, which will take responsibility of CH, when the energy level of CH drop down to a certain threshold value. The proposed algorithm works in two phases: Cluster setup phase and cluster steady phase. In first phase, clusters are formed and in the second phase, cluster head is enabled to receive the incoming packets, aggregate the data and report to the base station. This technique improves energy efficiency and cluster selection process.

Mantri et al. [42] proposed two tier cluster based data aggregation (TTCDA) mechanism in which functions based on additive and divisible aggregation are applied to data packets produced by every node by using temporal and spatial correlation. TTCDA works in three phases namely *cluster formation, intra-cluster aggregation* and *inter-cluster aggregation*. In the first phase, clusters are formed and cluster head is elected based on the highest available energy, number of nodes and distance to the sink node. As the name suggests, there are two levels of aggregation: First Local Aggregators (LA) perform intra cluster data aggregation, which forward the aggregated data to aggregator/gateway nodes (A/G).This aggregated data is then combined into one packet using division or additive functions depending upon the application requirements in tier 2. The packet generation rate of each node is usually different and not known at the network level. This technique is used to reduce packet count, computation and communication cost along with energy consumption and bandwidth. This approach can further be improved by considering mobility and heterogeneity of nodes.

Yue et al. [59] presented energy efficient and balanced cluster-based data aggregation algorithm (EEBCDA) in which one-hop communication is used to transfer the data from cluster head to base station. Network is divided into different size of rectangular regions called *swim lanes* and each swimlane is divided into number of rectangular regions called *grids*. This phase of network structuring is known as *Network Division phase*. A node having maximum energy is elected as cluster head in every grid. Cluster heads are rotated among the nodes in the grid, based on energy consumption. Grids which are further away from the base station are bigger in size and have more nodes to participate. In this approach, energy consumption is balanced, as if a CH which consumes more energy will have more sensor nodes to participate in cluster head election. This technique improves network lifetime and has balanced energy consumption but latency is more in distant grids.

Mantri et al. [46] proposed grouping nodes and clusters for efficient data aggregation (GCEDA) in which grouping of nodes is done based on available data. Cluster head (CH) uses divisible and additive data aggregation function to transfer the aggregated data to remote sink. Each cluster head determines the correlation factor among nodes which are 1-hop away. In proposed approach, the whole process of routing is divided into three phases. In first phase named *Cluster formation*, clusters are formed uniformly. Cluster head is selected based on Euclidean distance and highest energy. In *intra-cluster phase*, nodes with same data are found and grouped and then some aggregation function is applied depending upon the data. In *inter-cluster phase*, all cluster heads act as source nodes and send the aggregated data to the base station. Grouping of CHs helps to reduce energy consumption and improves network stability without considering heterogeneity of nodes and data accuracy.

Sinha and Lobiyal [60] presented energy efficient divergence based data aggregation technique. Algorithm works in two phases: *preliminary phase* and *final clustering phase*. In the first phase, the nodes sensing the same data are put in distinct cluster. In second phase, remaining sensors calculate their deviation from neighbor nodes and join the cluster with minimum divergence. In this approach, data aggregation is based on entropy of node. A window function is used to map the sensed data in the range of [0…1]. This technique is used to improve convergence rate, aggregation cycles, average packet drops, transmission cost and network lifetime without considering delay constraint. This approach can further be improved by considering heterogeneity of nodes and energy enriched multiple sinks. Rout and Ghosh [25] presented energy efficient adaptive data aggregation using network coding (ADANC). It is energy oriented cluster based data aggregation technique in which sensor nodes of a cluster are divided into two categories: *simple relay nodes* and *network coder nodes*. Network coder nodes act as *aggregator nodes*. Data aggregation is based on level of data correlation. If the level of data correlation factor is less in the received packets, then network encoding is performed. Otherwise traditional data aggregation is performed when the data correlation factor is high. Energy consumption is also reduced at node level with duty cycle. This technique is used to improve energy consumption by reducing communication cost, but data latency is not considered.

Xu et al. [61] proposed hierarchical data aggregation using compressive sensing (HDACS) based method to enable several compression thresholds dynamically based on different size of clusters at various levels of the data aggregation tree to reduce the data transfer amount. Instead of configuring only one node as sink, a hierarchy of multi-level cluster is configured for intermediate data collection. It reduces the data volume in data communication as compared to other compressive sensing approaches. DCT Based recovery algorithm is used for underlying domain. Further, cost is calculated based on both processor and radio energy consumption. The proposed method has been tested in SIDnet-SWANS simulation platform using real-world data and synthetic datasets.

Banerjee and Bhattacharyya [27] proposed evenly loaded distribution based data aggregation mechanism in which fuzzy logic is used to select the CH and distribute the equal amount of load among clusters to transfer data in optimal way. The main objective of this proposed approach is to address energy leakage problem of cluster heads which occurs due to uneven load distribution. The proposed approach is divided into rounds which are further divided into three phases namely *initial phase, regular setup phase* and *steady phase*. In initial setup phase, clusters are formed based on CSMA/CD technique. In regular setup phase, each node sends its energy status information to the cluster head which is further forwarded to the sink. The most powerful node becomes a new CH within a cluster and also weak zones are identified which can lead to disconnected zone in a region. To elect a cluster head, three parameters—energy level, distance to base station and distance from members of the cluster are considered. 'Mamdani' method has been used for fuzzification. In steady phase, a TDMA schedule is generated and broadcasted in a region by BS, so that each node can transmit its data only on its turn. This technique improves network lifetime but an overhead is incurred in regular setup phase for re-clustering and also latency is not addressed.

Mantri et al. [49] proposed bandwidth efficient cluster-based data aggregation (BECDA) technique for heterogeneous network by focusing on energy consumption. For applying the function of aggregation on the data generated by sensor nodes, this technique uses the concept of correlation of data within the packet. Random data is generated by each node between 0 and 1 using random function. There are three types of nodes having different energy levels: normal node (20 J), advance node (30 J) and super node (40 J). The variable amount of traffic is generated. A node with highest energy among the cluster members and a node with the highest number of neighbor nodes with one hop connectivity is considered as cluster head. This technique successfully improves bandwidth utilization, energy consumption and communication overhead. But it results in lesser throughput.

Chatterjea and Havinga [29] proposed diffusion and clustering based data-centric technique named CLUstered diffusion with Dynamic Data Aggregation (CLUDDA) which uses in-network processing to aggregate data. This approach combines directed diffusion with clustering approach. Rather than sending an *interest message* to each and every node (as in Directed Diffusion), it is communicated only to cluster heads. A new format for interest message is defined, in which query interpretation information along with query is sent. This mechanism is used so that a node can interpret the varying structure query of unfamiliar environment. There will be no need of fixed structure database storage. It also supports dynamic, layered data aggregation. The algorithm work in two phases: *interest propagation* and *data propagation*. It uses try and heal mechanism to restructure the network when some path fails. Data aggregation points also change as the location of required data sources change. This technique eliminates redundant processing and improves latency. But still, there is a huge memory requirement to store interest transformation and query responses. This approach is mostly dependent upon the underlying naming scheme.

Cluster based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 9. *Adaptive* data aggregation is performed at sink node in which aggregation ratio at CHs controls degree of spatial aggregation and reporting frequency at sensor nodes controls degree of temporal aggregation. In *distribution* based data aggregation, the goal is to distribute the equal amount of load among clusters to

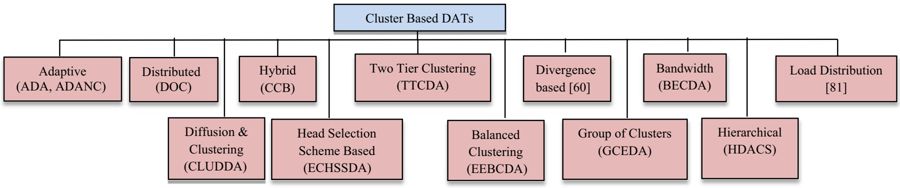


Fig. 9 Cluster based DATs taxonomy

transfer data in optimal way with minimum energy consumption. *Hybrid* based data aggregation also called combined data aggregation is used to enable dynamic aggregation by applying various techniques of clustering simultaneously and to select an adequate clustering technique based on status of network. *Head selection scheme* in data aggregation is used for selection of cluster head and formation of cluster. In *Two tier cluster* based data aggregation, functions based on additive and divisible aggregation are applied to data packets produced by every node by using temporal and spatial correlation at two levels. In *Balanced clustering* based data aggregation, one-hop communication is used to transfer the data from cluster head to base station. In *grouping nodes* based data aggregation, grouping of nodes is done based on available data. Cluster head (CH) uses divisible and additive data aggregation function to transfer the aggregated data to remote sink. In *energy* based data aggregation, sensors which are sensing the similar data are kept in different clusters while remaining sensors calculate their deviation from neighbor sensors and join cluster with minimum divergence and energy consumption. In *loaded distribution* based data aggregation, fuzzy logic is used to select the CH and distribute the equal amount of load among clusters to transfer data in optimal way with maximum network lifetime. The main objective is to mitigate energy leak problem of cluster heads. In *bandwidth efficient* cluster-based data aggregation, the goal is to maximize bandwidth utilization. Cluster head is elected based on the connectivity with 1-hop neighbors and energy level so that bandwidth which is wasted in case of multi-hop communication can be reduced. *Diffusion and clustering* based data-centric aggregation uses in-network processing to aggregate data and combines directed diffusion with clustering approach. In *hierarchical* data aggregation method, the main objective is to enable several compression thresholds dynamically based on different size of clusters at various levels of the data aggregation tree to improve the data transfer amount.

5.2.3 Concealed Data Aggregation Techniques

Concealed data aggregation is proposed to provide end to end encryption and in-network processing. Concealed data aggregation research work has been done by following authors. Ozdemir and Xiao [62] proposed integrity protecting hierarchical concealed data aggregation (IPHCDA) technique which uses holomorphic encryption based on elliptic curve cryptography to perform data aggregation in encrypted manner. This approach provides hierarchical data aggregation of encrypted data. It also provides integrity and confidentiality to the aggregated data by using message authentication codes (MAC). Base station uses encryption keys to classify the aggregated and encrypted data. In this technique, region of interest is divided into number of sub-regions, where each sub-region has its own public key for encryption. While aggregating the data in hierarchical manner, the calculated value of MAC is combined using XOR function. Then this encrypted aggregated data is forwarded to the base station. Base Station decrypts this data using Public key and MAC and assures data integrity. This technique improves trustworthiness, data confidentiality and integrity but does not consider latency.

Chen et al. [41] presented recoverable concealed data aggregation (RCDA) mechanism in which homomorphism encryption is used. It mitigates two problems of traditional encryption system with clustering. First, if cluster heads combine the encrypted data and forward that data to the base station, then base station will not able to estimate the maximum value of all sensing data. Secondly, base station cannot assure integrity by attaching signature to the sensed data samples. This technique uses the concept of recoverable (base station can recover data even after aggregation) aggregation. There are two variants of RCDA namely RCDA-HOMO and RCDA-HETE for homogeneous and heterogeneous types of networks. RCDA-HOMO works in four phases: *Setup, Encrypt-Sign, Aggregate* and *Verify*. In Setup phase, all the required secrets are prepared and installed for the BS. Encrypt-Sign is performed when a node sends the data to cluster head and aggregates the data in the next phase. In the last phase, BS verifies the authenticity using signatures. RCDA-HETE has five phases: *Setup*, *Intra-cluster Encrypt*, *Inter-cluster Encrypt*, *Aggregate* and *Verify*. In Setup phase, secrets are installed in high end and low end sensors. In Intra-cluster phase low end sensors send their sensed data to the high end nodes. In Inter-cluster phase, high end nodes collect and encrypt the collected data and put signatures for integrity. Integrity and confidentiality is verified by the base station in the last phase after verification. The proposed approach reduces communication cost and transmission overhead.

Lin et al. [63] proposed holomorphic public encryption system based concealed data aggregation technique i.e. CDAMA for multi-application environment in which base station collect data from aggregated cipher texts for particular application. This technique reduces the impact of compromising attacks and unauthorized aggregations. It also provides the feature of secure counting. Number of messages which are send for aggregation are known to the base station, so repeated and selective packet attacks are detected. The proposed technique improves robustness, reliability and energy cost which does not consider latency.

Zhang et al. [64] proposed slicing and mixing technology based balanced privacy-preserving data aggregation (BPDA) model which enhances the privacy-preserving efficacy by ensuring that slice can be sent to the nodes which have lower privacy preservation. The proposed model works in three phases: *Preparing phase, Privacy preserving* and *Data aggregation*. In first phase, an aggregation tree is prepared by following the standard approach i.e. TAG. In this phase, degree of each node is broadcasted by the node itself. Next the number of slices and threshold value for features is decided. Data features are recorded for further analysis of security efficacy. In the privacy preserving phase, nodes are selected for preservation and balance slicing is implemented. In the final phase, data is collected and aggregated. Node degree and energy are considered as QoS parameters and performance of proposed technique has been evaluated in terms of communication overhead and privacy-preserving efficacy.

Sicari et al. [65] presented discrete-time control loop based privacy aware dynamic data aggregation mechanism. It uses a private function dynamic data aggregation with privacy (DyDAP) functions to provide end to end secure data aggregation. Linear discrete time control theory is the underlying mechanism being used. The elements which are part of the proposed algorithm are represented by UML diagram. This technique reduces the communication load and improves congestion control, data accuracy, anonymity management but does not consider power consumption. Liu et al. [51] proposed an enhanced trust management method for data aggregation based on the strength of trust between sensor nodes and base station. Strength of ties between the nodes is calculated from the second hand information coming from other nodes using watchdog mechanism. The main objective of this model is to increase the level of security for data aggregation. The main objective is to classify the nodes as compromised or trustworthy nodes. It improves accuracy of data aggregation as compared to iRDTEDA protocol.

Concealed Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 10. In *hierarchical* concealed data aggregation, holomorphic encryption based on elliptic curve cryptography is used to perform data aggregation in encrypted manner. In *recoverable concealed data aggregation*, base station can easily estimate the maximum value of all sensing data and provide data integrity through the use of signatures. In holomorphic public encryption system based concealed data aggregation, base station collect data from aggregated cipher texts for particular application for *multi-application* environment. *Trust management* method for data aggregation is based on the connection between sensor nodes and base station. Balance *privacy-preserving* data aggregation enhances the privacy-preserving efficacy by ensuring that slice can be sent to the nodes which have lower privacy preservation using slicing and mixing technology. *Privacy aware* dynamic data aggregation provides end to end secure data aggregation using private encryption function.

5.2.4 Energy Based Data Aggregation Techniques

Data aggregation research work based on energy has been done by following authors. Çam et al. [32] proposed energy efficient secure pattern based data aggregation (ESPDA) mechanism to reduce transfer of duplicate data from node to cluster head. If similar data is sensed by sensor nodes then this technique produces pattern code to represent the characteristics of sensed data which are different. The task of cluster head is just to evaluate the pattern codes received and to select the nodes for data transmission. CHs use pattern codes to transfer data from node to base station without using encryption/decryption key distribution. It uses NOVSF Block Hopping technique to enhance security. This technique is used to improve occupied bandwidth rate and energy consumption. Feature extraction methods can further be incorporated for better pattern generation. Chen et al. [66] presented energy-efficient protocol for aggregator selection (EPAS) protocol (single level of aggregation) which is further extended to hierarchical EPAS (multiple levels of aggregation). The optimal number of aggregators for data aggregation process is estimated using power-consumption and generalized compression models. This technique improves power consumption and network lifetime. Mobile aggregators and selection of aggregators are two criteria which can be considered for future work.

Xu et al. [67] proposed cooperative communication based data aggregation mechanism whose main objective is to solve the traditional problem of CDA which is a NP hard problem. To solve this problem, a heuristic based approach MCT is used. In cooperative communication, a node is deployed with single transceiver and multiple source nodes can

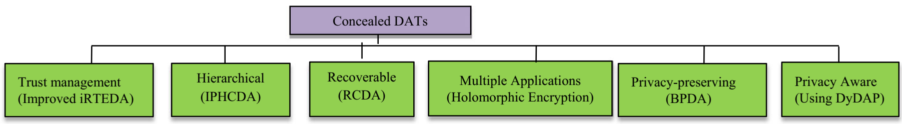


Fig. 10 Concealed DATs taxonomy

send their data to that node. DMCT is the variant of MCT for distributed environment implementation. This technique minimizes energy cost but incurs a lot of computational cost. Li et al. [36] proposed distributed and latency based data aggregation technique to prevent conflicts among neighboring clusters during transfer. The well-known minimum-latency aggregation schedule (MLAS) problem has been addressed and an energy-efficient distributed scheduling algorithm i.e. Clu-DDAS based on a novel cluster-based aggregation tree has been proposed. It constructs cluster based tree and their respective cluster heads for efficient transfer of data among clusters. First of all, cluster-based data aggregation tree (CluDAT) is constructed, to provide collision free communication among clusters. The proposed algorithm has a latency bound of $4R + 2\Delta - 2$, where Δ is the maximum degree and R is the interior network radius and is generally smaller than the network radius R. The proposed approach improves transmission cost, latency and network lifetime without achieving desired level of energy consumption.

Xiang et al. [68] presented greedy heuristic and mixed integer programming formulation based compressed data aggregation technique which considers the concept of compressed aggregation and joint routing. Further this technique obtains optimal aggregation trees to solve both large and small scale problems. The problem of minimum energy compressed data aggregation (MECDA) has been defined and the characterizations of the optimal solutions are provided. A nontrivial mixed-integer programming (MIP) formulation of MECDA has been discussed, through which the optimal solutions for small scale WSNs has been addressed. A greedy heuristic has been proposed to deliver near optimal solutions for large scale WSNs. It improves communication cost but computation cost has not been considered.

Li et al. [16] presented data accuracy based an Energy-Efficient and High-Accuracy (EEHA) scheme for secure data aggregation. This technique provides high accuracy without releasing readings of private sector and reduces overhead on sensors (battery limited). There are four phases of the proposed technique i.e., *aggregation tree construction, slicing, mixing* and *aggregation.* In the first phase, an aggregation tree is constructed by simply taking the union of all paths available in a region or some traditional approach such as TAG is used. In the slicing phase, a leaf node randomly selects the K number of nodes within h hops. Then that particular node divides its data packet into K units and encryption is performed on those slices. In mixing phase, a node waits for a certain period of time to receive slices of the neighbor nodes. A public key is shared with the sender for decryption purpose. After getting all the slices from the neighbor nodes, a node will sum up all the data received, encrypts it and sends to the base station in the last phase. This technique is designed to improve security and data accuracy with larger energy consumption. Kuo and Tsai [43] proposed tree based data aggregation mechanism by using relay nodes to construct a data aggregation tree to reduce power consumption of data transmission. Tree construction with or without relay node is considered as a NP- complete problem. Authors proposed a two approximation algorithm based on shortest path mechanism in a distributed fashion. This technique improves energy consumption but network lifetime and communication overhead are not considered.

Liu et al. [17] proposed High Energy-Efficient and Privacy-Preserving (HEEPP) data aggregation technique. There are six phases in which the proposed technique works: *formation of aggregation tree, slicing, provides security through encryption and decryption while transfer, mixing and resembling, aggregation of data* and *query mechanism* to process data. In first phase, aggregation trees are formed and a hierarchy is established using slicing technique. A parameter MAX_CHILDREN represents the maximum number of children node a parent can have. Initially a base station will send the HELLO packet to all nodes. Those nodes which do not have parent nodes send a JOIN request to the BS. In second phase, nodes wait for a certain period of time for all the packets to receive. All received packets are checked to get shared keys and to be prepared for decryption. To encrypt a particular message, shared key is used and then it is sent to the neighbor nodes. In the last phase, data is aggregated. This technique improves aggregation accuracy, data integrity, computation overhead and communication overhead with huge energy consumption.

Chao and Hsiao [69] proposed structure free and energy-balanced (SFEB) based data aggregation technique. It uses two phase data aggregation along with dynamic aggregator selection scheme. The proposed technique works in two phases. In first phase selection of dynamic aggregators is performed to collect the data from their respective regions. In phase two, data is aggregated and reported to the higher level node. A lightweight aggregator election algorithm is designed to avoid costly data exchange rounds. To improve aggregation ratio, a blend of data-aware anycast (DAA) and randomized waiting (RW) is followed. It improves aggregation ratio, maintenance overhead and energy consumption. Balanced energy depletion is still an issue to overcome. Engouang et al. [70] presented game based data aggregation mechanism by considering ZigBee and Self-Fulfilling Belief (SFB) for security and two player's game theory. Tree and star based topologies are used to improve energy consumption. The main objective is to provide collision free communication to the nodes. Claude Shannon's information theory is considered for maintaining privacy. Based on SFB, every player plays its best with its energy parameter and also assumes that other players are also trying their best to win the game. This technique improves trustworthiness, power consumption, end-to-end delay and throughput but network lifetime and communication overhead are not considered. Liu et al. [71] proposed reliability based data aggregation technique by using iRTEDA protocol which combines residual energy, reputation system, recovery mechanism and link availability to improve security. Based on a Beta distribution, a reputation function is used to calculate trustworthiness and reputation of nodes. When an aggregator node is identified as compromised node, then reselection of aggregator node is initiated based on residual energy level and availability of link. This technique improves robustness, network lifetime, energy by excluding compromised nodes, data accuracy without considering latency.

Pourpeighambar et al. [2] presented data aggregation for mobile object using rate distortion (RD) with static clustering (DAMORD-SC) in which predefined clusters are divided into grids and perform data aggregation through pre-calculated correlation matrices. It is assumed that there is a moving object in the field which also generates the data. The main objective is to reduce the data redundancy due to spatial correlation using rate distortion (RD) theory. In this approach, each cluster head sends an advertisement message to the network. Nodes estimate their distance with respect to the cluster heads and join the cluster with the minimum distance. Cluster head sends a TDMA schedule to all those nodes whose join requests are received. A target tracking component is installed in each node and each node is aware of location information. Firstly, all cluster members are in sleep mode. All nodes awake simultaneously to track the position of object and send the data to the cluster head. At the end of each round, last node is responsible for sending the number of grids in which the object resides along with its own data. Data accuracy and energy consumption parameters are used to create the adequate grid size of clusters to improve energy consumption and data accuracy but delay is larger. Tsai et al. [20] proposed delivery delay based data aggregation technique by using the concept of buffered packets and configure the parameters in machine-to-machine (M2M) networks. A periodic per-hop timing control method is deployed in which if the amount of data that is to be buffered increases or buffering time is taking too long then aggregation rate can be increased to reduce the energy consumption. If buffering time is implemented with little variance, it can significantly improve the energy consumption in both line based and grid based scenario. This technique improves latency and network lifetime without considering data accuracy.

Krishna and Doja [72] proposed Multi-Objective Meta-Heuristic approach for Energy-Efficient secure Data Aggregation(MH-EESDA).This technique uses divide-and-conquer technique to create the secure clusters. Proposed technique works in three consecutive steps: *form clusters, select secure nodes* and a*cross the secure route paths, energy efficient data aggregation is performed.* Performance of proposed technique has been evaluated in simulated environment in terms of threshold-based degree of intrusions. Ramachandran and Porter [73] proposed a remote component binding technique named Hitch Hiker to offer sustainance for multi-hop data aggregation. The proposed model is based on priority information associated with bindings. There are two types of bindings namely high-priority bindings and low-priority bindings. For radio transmission component high priority remote binding is defined and for those communicate using data aggregation overlay, low priority binding is defined for non-critical applications. A component meta-data is used to determine bindings of remote component and to create a network of multi-hop overlay within the payload's unused space of current flows of traffic. Further, evaluation of proposed technique has been conducted in simulated environment which shows that proposed technique consumes fewer resources and reduces energy consumption.

Lee et al. [50] proposed restructuring binomial trees and energy-efficient delay-aware data aggregation by considering energy consumption scenario like data transmission between sensor node and cluster head. Further, AVL trees and red–black trees are used to make tree balanced by using the rotation method. Bionomical tree is rebuilt using this rotation mechanism, in which roles are swapped by sensor nodes periodically to keep the network balanced. Delay-aware data collection (DADC) and delay-minimized energy-efficient data aggregation (DEDC) are used for the initial construction of the tree. DADC is efficient in terms of latency and constructs the tree in bottom up fashion. Information like node ID and location are exchanged in the scheme. A connect message is send by a node to its neighboring node so that they can form a convergence cluster and this process continues until a binomial tree is generated. Both data aggregation delay and energy efficiency of the sensor nodes has been improved which reduces aggregation cost.

Xiao et al. [74] presented data aggregation technique namely centralized energy allocation (CEA) based on the immune-genetic heuristic. The main objective of proposed technique is to find the effective strategy of energy allocation to maximize the precision of the data aggregated received by the sink. Immune-Genetic heuristic is the result of artificial immune algorithm (AIA) and genetic algorithm blend. It provides better optimal solutions as compared to these approaches and provides faster convergence performance. The underlying model of proposed approach assumes the link failure factor for considering unreliability in incoming and outgoing links. Based on Gibbs sample, a localized algorithm is proposed which can adapt itself to the dynamic distributed WSNs. Further, this technique discusses the tradeoff between energy consumption and data quality of network with heterogeneous nodes. Zhang et al. [75] presented energy based data aggregation technique namely data aggregation supported by dynamic routing (DASDR) in which two potential fields has been considered: depth potential (to ensure that packets are reaching the sink at least) and queue potential (to create packets more spatially convergent). DASDR utilizes only local information for routing decisions based on potential dynamic routing mechanism. A queue potential field is implemented to make packets spatially convergent by exploring local information of queue length. Due to which it can rapidly forward the packets further in less amount of energy in mobile and static scenarios. This technique improves energy consumption but does not perform well in heavy traffic scenario as expected.

Energy based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 11. Energy efficient *secure pattern* based data aggregation is used to reduce duplicate data transmission from node to cluster head. *Hierarchical* energy-efficient protocol for aggregator selection protocol is used to perform multiple levels of aggregation. *Co-operative* communication based data aggregation is using a heuristic based approach MCT, in which a node is deployed with single transceiver and multiple source nodes can send their data to that node. *Distributed and latency based data aggregation* is used to prevent conflicts among neighboring clusters during transfer. Greedy heuristic and mixed integer programming formulation based *compressed* data aggregation technique considers the concept of compressed aggregation and joint routing. *Data accuracy* based secure data aggregation technique performs aggregation of data without releasing readings of private sensor and reduces overhead of sensors (battery limited). *Tree based* data aggregation uses relay nodes to construct a data aggregation tree to reduce power consumption of data transmission. *Privacy based* data aggregation performs six steps sequentially: formation of aggregation tree, slicing, provides security through encryption and decryption while transfer, mixing and resembling, aggregation of data and query mechanism to process data. *Structure free* data aggregation uses dynamic aggregator selection process and two-phase aggregation mechanism to improve aggregation ratio, maintenance overhead and energy consumption. *Game* based data aggregation considers ZigBee and self-fulfilling belief (SFB) for security and two player's game theory along with tree and star based topology is used to improve energy consumption. *Reliability* based data aggregation technique uses iRTEDA protocol which combines residual energy, reputation system, recovery mechanism and link availability to improve security. *Static clustering* based data aggregation divides predefined clusters into grids and performs data aggregation through pre-calculated correlation matrices. *Delivery delay* based data aggregation technique uses the concept of buffered packets and configures the parameters in M2M networks for data aggregation. Energy-efficient *multi-objective meta-heuristic* is used for secure data aggregation using divide-and-conquer technique to create the secure clusters. *Remote component binding* data aggregation offers sustainance for multi-hop data aggregation, in which component meta-data is used to determine bindings of remote component and to create a network of multi-hop overlay within the payload's unused space of current flows of traffic. Restructuring binomial trees and energy-efficient *delay-aware* data aggregation considers energy consumption scenario like data transmission between sensor node and cluster head. Data aggregation uses centralized algorithm based on the

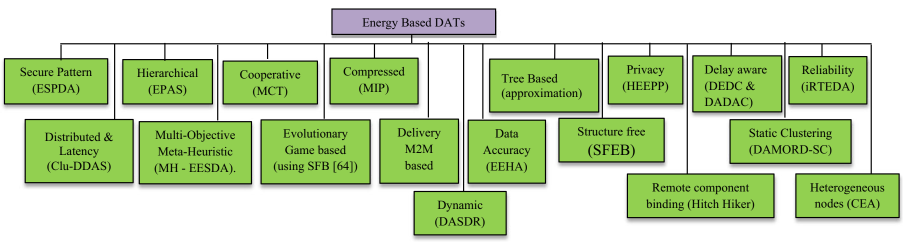


Fig. 11 Energy based DATs taxonomy

immune-genetic heuristic to find the effective strategy of energy allocation to maximize the precision of the data aggregated which is received by the sink in *heterogeneous network*. In *dynamic* data aggregation, two potential fields have been considered: depth potential (to ensure that packets are reaching the sink at least) and queue potential (to create packets more spatially convergent).

5.2.5 Latency Based Data Aggregation Techniques

Data aggregation research work based on latency has been done by following authors. Lu et al. [14] proposed latency based energy efficient adaptive data aggregation DMAC technique which has three advantages over SMAC which are improved packet delivery latency, no more fixed duty cycle and less collision due to fixed synchronous duty cycle. Before this, all other MAC protocols suffer from the problem of *Data Forwarding Interruption problem*, which arises due to the nodes residing in sleep mode to which data is forwarded next. But this problem is removed in proposed DMAC, as it follows staggered active/sleep schedule, so that data can be forwarded in continuous flow in a multi-hop fashion. It uses data prediction mechanism along with *More -To-Send* packet, by which a parent node in a tree structure will get to know that whether its child node has more data to send or not. This approach ensures reduced energy consumption and improved latency, but DMAC may not be suitable in a burst traffic scenario.

Li et al. [18] presented distributed data aggregation algorithm namely cell-aggregation scheduling (Cell-AS) which uses physical interference model. It uses O(K) time slots to finish the task of aggregation, where k is logarithm of ratio between the shortest and longest link length in WSN. In this approach, network is divided into cells according to the value of k to alleviate the need of global information. Using the concept of nearest-neighbor criterion, minimum latency aggregation schedule (MLAS) problem is addressed under physical model to reduce aggregation latency, transmission cost and improve network lifetime. Link scheduling, power control and joint tree construction are key points of the proposed model. Energy cost at local node level can be further improved. Liu et al. [76] proposed approximation based conflict-aware data aggregation technique to reduce latency. It uses two models of directional antenna (switch beam and steering beam) along with protocol interference model. In the proposed algorithm, the latency of schedule is close to depth of BFS approach (which is the lower bound of latency). Nodes away from sink need a schedule for data aggregation. In switch beam directional model, sensors have directional sending antennas and Omni-directional receiving antennas. This technique performs extremely well and reduces schedule latency.

Latency based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 12. Energy efficient *adaptive* data aggregation uses DMAC to deliver data in unidirectional trees for data aggregation. *Distributed* data aggregation uses physical interference model which uses O(K) time slots to finish the task of aggregation, where *k* is logarithm of ratio between the shortest and longest link length in

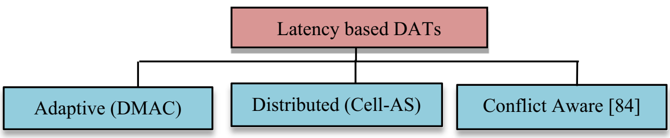


Fig. 12 Latency based DATs taxonomy

WSN. Approximation based *conflict-aware* data aggregation reduces latency and uses two models of directional antenna (switch beam and steering beam) along with physical interference model.

5.2.6 Network Lifetime Based Data Aggregation Techniques

Data aggregation research work based on network lifetime has been done by following authors.

Kalpakis et al. [13] proposed energy based data aggregation technique in which energy-constrained sensors are deployed to monitor vicinity periodically. A polynomial time, near optimal maximum lifetime data aggregation (MLDA) algorithm is proposed. Data is gathered from all the sensors systematically and is transferred to base station through in-network aggregation. Energy cost and latency is improved but network lifetime is larger. Xue et al. [77] proposed linear programming based data aggregation mechanism. In this approach a data aggregation problem is considered as multi-commodity flow problem, where each source node is taken as a commodity. An approximation algorithm is proposed which reduces the running time of the optimal solution by factor N, where N is the number of commodities assumed. It utilizes tree based data aggregation by setting up of multiple base stations to reduce the running time but energy consumption and communication costs are larger. Tang and Xu [78] presented precision constrained based data aggregation technique. There are three factors which effect the lifetime of a network: Residual energy of node, varying pattern of sensor readings and communication cost between the sink and nodes. Precisions can be differentiated by partitioning the precision constraint and allocating the error bounds (dynamically adjustable) to every sensor node to improve energy efficiency and latency. It has been concluded that Uniform precision allocation does not perform good due to distributed nature of WSNs even if the patterns sensed are similar in nature. Balancing the energy depletion of network is more important than reducing the energy consumption at node level.

Yum and Hua [79] presented distributed data aggregation technique using smoothing approximation function and underlying routing scheme to design distributed gradient algorithm. The proposed approach combines data aggregation with maximum lifetime optimized routing. Network lifetime can be extended in two ways. One method is to reduce the traffic across the network, so that communication cost can be reduced in terms of energy. Second way is to apply data aggregation to compress the traffic data generated by nodes. Geometric routing is used along with correlation model. This technique reduces the data traffic but network lifetime and communication overhead are not considered. Shan et al. [19] presented the problem of finding a shortest path tree while applying data aggregation with the maximum lifetime. A distributed data aggregation technique is used in which a problem is transformed into a tree in a polynomial time. The addressed problem is solved by load balancing approach at every level of fat tree. This technique improves network lifetime but delay is larger. Azad and Sharma [52] presented Pareto-optimal theory based multi-criterion decision-making approach for effective data aggregation technique. The selection of cluster heads is based on energy consumption and cluster density. A region is covered by multi objective-based decision-making problems and there is possibility of one of the three solutions. First, the solution is completely dominated. Second, the solution is neither dominated nor dominating and third, solution is non-dominated. Dominating means that one solution overshadows the properties of alternative solutions. In the proposed technique, TOPSIS which is one of the multi objective approaches are used. This technique performed better in terms of network lifetime as compared to distributed hierarchical agglomerative clustering based data aggregation.

Asemani and Esnaashari [53] proposed an algorithm called Learning Automata based aGgregation (LAG) which is used to find the route to transfer data after aggregation. It is capable of adapting itself dynamically in changing environment and to choose innovative paths towards the sink consequently. Every node is trained with learning automata (LA) which aids the sensor node to choose its next hop for transferring data towards the sink. INCASE-LA, learning automata is used in the proposed technique. Performance of proposed model has been evaluated in simulated environment which clearly shows that proposed model performs better in terms of power consumption and network lifetime. Awang and Agarwal [80] proposed data aggregation technique where aggregator nodes are determined speculatively without being dependent on global knowledge i.e. nodes' geographical location, network topology and data flow. Received signal strength indicator (RSSI) is used to aggregate and route data packets. Each data packet generated by a node is identified by an aggregation ID. Nodes which are closer to the aggregator node are favored for data forwarding process. Performance of data aggregation technique has been evaluated in terms of network lifetime, end-to-end delay, average energy consumed per data packet and packet loss probability.

Network Lifetime based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 13. In *Energy* based data aggregation, energy-constrained sensors are deployed to monitor vicinity periodically. *Linear programming* based data aggregation uses tree based data aggregation by setting of multiple base stations to reduce the running time but energy consumption and communication costs are larger.

In *Precision constrained* based data aggregation, precisions can be differentiated by partitioning the precision constraint and allocating the error bounds (dynamically adjustable) to every sensor node. *Distributed* data aggregation uses smoothing approximation function and underlying routing scheme to design distributed gradient algorithm. *Shortest path* based data aggregation uses shortest path aggregation trees to develop a distributed protocol which improves network lifetime. Pareto-optimal theory based *multi-criterion decision-making* data aggregation is used for the selection of cluster heads by considering energy consumption and cluster density as parameters. *Learning automata based data aggregation* is used to find the route to transfer data after aggregation and is able to adapt itself dynamically in changing environment and to choose innovative paths towards the sink consequently. In *end-to-end delay* based data aggregation, aggregator nodes are determined speculatively without being dependent on global knowledge i.e. nodes' geographical location, network topology and data flow.

5.2.7 Nature-Inspired Optimized Data Aggregation Techniques

Data aggregation research work based on nature-inspired optimized has been done by following authors. Misra and Mandal [81] proposed search space based data aggregation

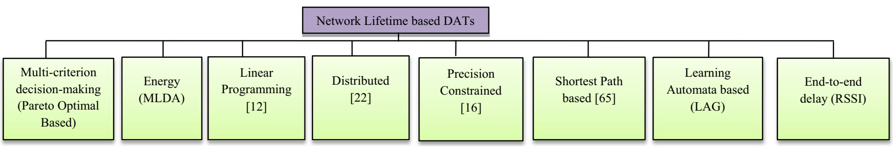


Fig. 13 Network lifetime based DATs taxonomy

technique which considers population-based algorithm (Ant colony system) to design aggregation tree iteratively. Node Set act as an input to this algorithm and ants are assigned to the nodes. Routes are searched by ants and are communicated via pheromones. Node potential is considered as an estimate of distance to the final destination. Each ant iterates until an aggregation tree is constructed and converges to local optimal solution. The main goal is to improve energy consumption, transmission cost and network lifetime but communication overhead is not considered.

Yucheng and Fan [82] presented energy based data aggregation technique in which data aggregation tree is constructed to transfer sensed data to single sink node using the path designed by pheromone in ant colony algorithm. In first phase, next hop node is selected. In second phase, a path is created around the node which is just selected. In third phase, pheromone value of neighboring nodes is updated. This technique reduces energy consumption but data accuracy is not considered. Lin et al. [83] presented network lifetime based data aggregation ant colony algorithms (DAACA) technique which considers three phases i.e. *initialization, packets transfer* and *operations on pheromones*. The value of energy and amount of pheromone of neighboring nodes can be estimated to dynamically select the next hop for transmission. Based on local and global information, pheromones are adjusted after certain number of transmission rounds. Routing tables are maintained by the nodes themselves. This technique is used to improve energy efficiency, computation complexity and success ratio but network lifetime and communication overhead are not considered. Ho et al. [84] presented ladder diffusion and ACO based data aggregation technique to find the path for transmission and data relay while preventing the generation of circle routes. Back-up routes are also maintained in case of path failure. Energy dissipation and processing time are reduced without considering data accuracy.

Lu et al. [85] proposed semi-structured multi- objective tree based dynamic data aggregation protocol in which ant colony optimization (ACO) is used to find the optimal route for data transfer. Further, concept of sliding window is used for prediction of arriving packets to improve the aggregation probability and reduce the transmission delay. For data aggregation, this protocol also shows that convergence of packet transmission from both temporal and spatial point of view. To achieve spatial convergence, data-aware anycast (DAA) is proposed and randomized waiting (RW) is preferred for temporal convergence. Finally, simulation results validate the feasibility and the high efficiency of the novel protocol when compared with other existing approaches. Paul and Gopinathan [86] proposed ACO based hybrid data aggregation mechanism which aggregates fruitful message through shortest path. Anomaly messages are identified and removed by a classification technique based on support vector machine. First of all, classification is done based on geographical location and the most powerful node is elected on the basis of computational capacity. At the cluster level, ant colony is implemented in order to collect the data from other nodes. Then messages are classified as anomaly or fruitful messages. Fruitful messages are forwarded further to base station for future decisions. This technique improves network life time but classification and communication cost is larger.

Nature-inspired Optimized based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 14. *Search space* based data aggregation considers population-based algorithm (Ant colony system) to design aggregation tree iteratively in WSN to improve energy consumption, transmission cost and network lifetime. In *Energy* based data aggregation, data aggregation tree is constructed to transfer sensed data to single sink node using the path designed by pheromone in ant colony algorithm. *Network lifetime* based data aggregation considers three phases namely *initialization, packets transfer* and *operations on pheromones*. *Ladder diffusion* based data

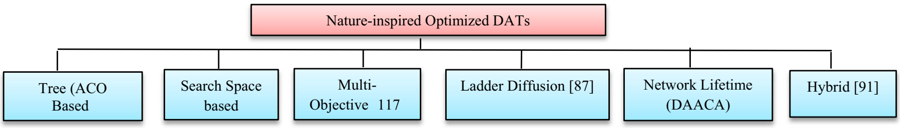


Fig. 14 Nature-inspired optimized based DATs taxonomy

aggregation is used to find the path for transmission and data relay while preventing the generation of circle routes to avoid energy wastage and processing time. In ACO based *hybrid data aggregation*, support vector machine is used for the removal of anomaly message after aggregation of fruitful message through shortest path. *Tree based data aggregation* is used to find the optimal route for data transfer using ACO in which prediction of arriving packets is done.

5.2.8 QoS Based Data Aggregation Techniques

Data aggregation research work based on QoS has been done by following authors. Pham et al. [87] proposed a new metric to assess the quality of data aggregation process i.e. data aggregation quality (DAQ). It is energy based data aggregation metric which do not require prior knowledge of statistical distributions of sensing data. Enhanced LEACH and clustered PEGASIS protocols are proposed, in which it has been observed that chain based algorithms are more efficient in terms of energy but not in terms of DAQ. Example is chain based PEGASIS, which performs well in terms of energy. Some event driven models can be used for simulation purpose in future.

Li et al. [15] proposed Markovian chain based waterfalls random partial aggregation technique to analyze the Trade-Off Index (TOI) of different applications. In non-aggregation applications, a large amount of energy is consumed. In aggregation based algorithms, latency is the problem. So to alleviate the problems of two approaches, partial aggregation is proposed along with TOI. By TOI, we can observe the tradeoffs among data accuracy, delay and energy depending upon the application's requirement. But only simple network scenarios are considered for simulation in the proposed work. Chen et al. [88] proposed adaptive fault tolerance based data aggregation technique in which a mathematical model is designed to fulfill QoS requirements of applications. There is a tradeoff between reliability and energy consumption. To provide reliability or fault tolerance there is a need for redundancy of source or paths. But this also leads to energy depletion. Author proposed a model to find the optimal level of redundancy so that reliability can be ensured without considerable energy consumption. Noise model can be considered in future work, so that more realistic scenario can be analyzed.

Misra and Thomasinous [89] presented LEO: simple least-time energy-efficient routing protocol with one-level data aggregation. It is time and energy based data aggregation mechanism in which shortest path is used to transfer data from source to destination. Each node stores information about its neighbor only to save memory. It increases network lifetime as task of data aggregation is performed by next node of source node. So in this role of data aggregation is shuffled each time when data transfer takes place. The estimated time to reach the BS from that particular node and residual energy information is maintained in the neighbor node table of each node. This technique improves latency, average energy consumption and average network lifetime. Mobility, security and load balancing are some key features which can be included to make this approach versatile. Chen et al.

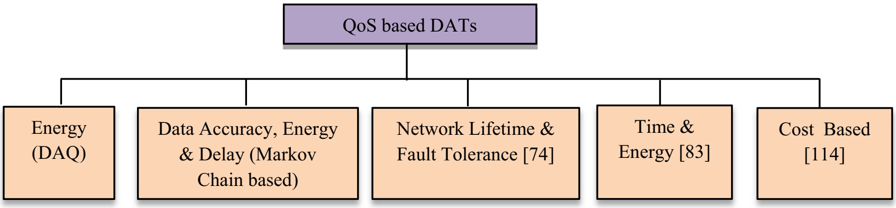


Fig. 15 QoS based DATs taxonomy

[90] proposed cost based data aggregation technique using in-network aggregation in which data is aggregated at the intermediate nodes from route to the sink using two different functions: (1) summation functions (sum, mean, and weighted sum) and (2) extreme functions(max and min). Further, an algorithm has been presented to find the tradeoff between finding a low cost path to the sink and local aggregation of flows in terms of cost taking as a QoS parameter.

QoS based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 15. In *energy* based data aggregation, quality of data aggregation process is accessed using data aggregation quality (DAQ) metric which does not require prior knowledge of statistical distributions of sensing data. In *data accuracy, energy & delay* based data aggregation, Markovian chain based waterfalls random partial aggregation is used to analyze the Trade Off Index (TOI) of different applications. Adaptive *network lifetime* and *fault tolerance* based data aggregation uses mathematical model to fulfill QoS requirements of applications and to find the tradeoff between reliability and energy consumption. *Time and energy* based data aggregation uses shortest path to transfer data from source to destination and each node stores the information about its neighbor only to save memory. In *cost based data aggregation*, data is aggregated at the intermediate nodes from route to the sink using two different functions: (1) summation (sum, mean, and weighted sum) and (2) extreme (max and min).

5.2.9 Scheduling Based Data Aggregation Techniques

Data aggregation research work based on scheduling has been done by following authors. Hu et al. [31] presented timing control protocol based dynamic data aggregation technique to calculate the minimum aggregation period to reduce energy consumption and latency. According to the quality of data aggregation required, aggregation time period can be dynamically changed. An intelligent timer and high level knowledge of the network is considered. Data sink requests for certain number of reports with maximum latency period. If the number of reports is received before the maximum time, then the value of minimum aggregation time is changed. The proposed algorithm works in two phases: *setup phase* and *data collection phase*. In setup phase, tree formation message is broadcasted in the network. There are three types of aggregation which can be implemented: *full aggregation, per-level aggregation* and *no aggregation*. In data collection phase, each node will transmit its data to the intended receiver within a particular time period.

Yu et al. [34] proposed maximal independent sets based distributed data aggregation technique to produce collision-free schedule. It uses greedy strategy to improve time latency. The time bound given by the algorithm is given by $24D + 6\Delta + 16$ where D is the diameter of the network and Δ is the maximum node degree. The proposed algorithm works in two phases namely *Construction of distributed aggregation tree* and second is *distributed aggregation scheduling* using underlying MAC protocol for distributed nature. The underlying model can adapt itself and can schedule according to the network conditions as some node dies or joins the network. Bagaa et al. [91] proposed semi-structured and unstructured based data aggregation mechanism in which unstructured topology (DAS-UT) and semi-structured topology (DAS-UT) is used to construct and execute aggregation tree simultaneously. A node is having a number for options for parent selection and time slot for reuse. The proposed algorithm works in two phases: First phase is *network organization* phase and second phase is *simultaneous data aggregation tree construction and data scheduling*. In network organization phase, network is structures into layers based where each layer is having nodes with same number of hops. In DAT-ST, each node sends its data to the parent node only in its transmission slot. The proposed technique improves latency and network lifetime.

Li et al. [48] presented improved distributed data aggregation technique to generate a collision-free schedule. It is implemented by using breadth first search tree at sink node. The latency bound is given by $61R + 5\Delta - 76$, where R is the radius of the region with sink node as center, Δ is maximum node degree. It is an improved algorithm of approach proposed by Bo [34]. Jhumka et al. [92] proposed collision free data aggregation scheduling (DAS) in which sensor nodes aggregates data and transfers to a sink node. By using time division multiple access (TDMA) technique timeline is divided into time slots. DAS problem has been extensively studied and necessary condition to solve the same is identified. DAS problem has been presented as two variants namely strong DAS and weak DAS. It has been analyzed that strong DAS problem cannot be solved. Weak DAS also cannot be solved if there are crash losses in the system. This technique improves fault tolerance without considering data accuracy.

Joo et al. [93] presented delivery delay based data aggregation technique using node-exclusive interference model. There are two types of scheduling approaches considered namely *myopic* and *non-myopic*. In myopic scheduling, present scenario or state is considered for scheduling decisions to be taken. In Non-Myopic, future requirements are considered while scheduling. A scheduling algorithm is proposed to minimize the overall delay and to rationalize the gain of data fusion in tree topology networks. The features and characteristics of optimal scheduling have been analyzed so that lower bound of performance can be calculated. By estimating the next state of the present network, performance can be significantly improved. Non-myopic approach improves end to end delay, transmission cost and network lifetime but desired data accuracy is not achieved. Bagaa et al. [94] proposed distributed algorithm for integrated tree construction and data aggregation which considered data aggregation scheduling problem. A collision-free schedule is determined to find the route with minimum time duration to transfer aggregated data to the base station from sensor nodes. Authors proposed distributed algorithm for integrated tree construction (DICA) along with data aggregation for the tree formation. While tree formation, a node has multiple available options for the selection of the parent node. Further, proposed improves communication overhead. Finally, performance of an algorithm has been evaluated using TinyOS with distributed protocol.

Kwon et al. [95] presented timeout control based data aggregation mechanism which enables the changes in timeout dynamically as per currently accumulated data in a node. This technique avoids control over messages thus it is not only robust but also improves delivery delay and waiting time. Local information is utilized without including control messages. The main focus of the proposed schemes is on routing layer as compare to the traditional approaches in which MAC layer is considered significant. There are two

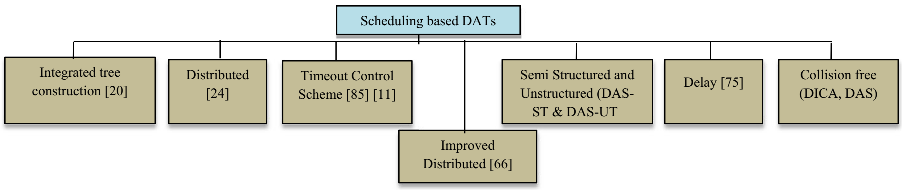


Fig. 16 Scheduling based DATs taxonomy

methods for implementing timeout control. First method is to set a temporal limit (in seconds, milliseconds etc.). Second method is to fix target of data of aggregation. The proposed scheme uses the first method for timeout control. Node's position and multi-hop behavior of nodes are yet to be considered.

Scheduling based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 16. Timing control protocol based *dynamic* data aggregation is used to calculate the minimum aggregation period to reduce energy consumption and latency. Maximal independent sets based *distributed* data aggregation is used to produce collision-free schedule and uses greedy strategy to improve time latency. *Semi-structured and unstructured* based data aggregation mechanism uses unstructured topology and semi-structured topology to construct and execute aggregation tree simultaneously and improves selection of parent node and time slot for reuse. *Improved distributed* data aggregation is used to generate a collision-free schedule by using breadth first search tree at sink node. In *collision free* data aggregation, sensor nodes aggregate data and transfer to a sink node by using TDMA, in which timeline is divided into time slots. *Delivery delay* based data aggregation uses node-exclusive interference model to perform aggregation of data by using two types of scheduling approaches considered namely myopic (considers present scenario or state for scheduling decisions to be taken) and non-myopic (future requirements are considered while scheduling). *Integrated tree construction* based data aggregation considers data aggregation scheduling problem, in which collision-free schedule is determined to find the route with minimum time duration to transfer aggregated data to the base station from sensor nodes. *Timeout control* based data aggregation enables changes in timeout period dynamically as per currently accumulated data in a node.

5.2.10 Tree Based Data Aggregation Techniques

Data aggregation research work based on tree has been done by following authors. Tan et al. [96] proposed energy based autonomic data aggregation mechanism named localized power-efficient data aggregation protocol (L-PEDAP) in which local minimum spanning tree (LMST) [29] and relative neighborhood graph (RNG) topologies are used to construct a minimum spanning tree. Based on MST, the position of next adjacent hop is calculated to find the shortest path between sender and receiver to increase the network lifetime. Whenever a route or node fails, route maintenance phase is initiated. This algorithms works well for certain threshold value of node density, beyond which it can results in poor performance. Costs of setup and maintenance are being ignored in the proposed model.

Virmani et al. [44] presented adaptive energy aware data aggregation tree (AEDT), in which a node with maximum energy is elected as an aggregator node. All other nodes except the sink node and aggregator node go to the sleep mode, so it reduces energy consumption. There is a threshold value for traffic, beyond which packets are selected according to the parent capacity. If the estimated traffic load is greater than the communication capacity of parent node then an OVERLOAD message is transmitted to the network. A memory table is also maintained in which all the paths are stored. After *t* seconds (where *t* is real time delay), proposed tree is refreshed and new root node (parent) is elected if required. Shortest path and energy are the two important factors which are considered while routing. But overhead of maintaining a memory table of such a dynamic network is an overhead. Rout and Ghosh [25] presented energy efficient adaptive data aggregation using network coding (ADANC). It is energy oriented cluster based data aggregation technique in which sensor nodes of a cluster are divided into two categories: *simple relay nodes* and *network coder nodes*. Network coder nodes act as *aggregator nodes*. Data aggregation is based on level of data correlation. If the level of data correlation factor is less in the received packets, then network encoding is performed. Otherwise traditional data aggregation is performed when the data correlation factor is high. Energy consumption is also reduced at node level with duty cycle. This technique is used to improve energy consumption by reducing communication cost and network lifetime but data latency is not considered.

Hakoura and Rabbat [97] presented collective tree protocol (CTP) based data aggregation mechanism which is compared with two gossip algorithms namely *Broadcast Gossip* and *Pairwise Randomized Gossip*. CTP collects and aggregates data using spanning tree which is an efficient method for data aggregation. In the initial phase, a beacon packet is broadcasted in a network by nodes containing unique ID of nodes. For discovering neighbor node and to check the quality of links, beacons are used. Topology update message is propagated in case of node or link failure. The performance is being measured in terms of number of transmissions for distributed averaging. But maintaining a minimum spanning tree incurs an overhead cost. Intanagonwiwat et al. [28] proposed tree based data aggregation technique in which greedy approach is used to adjust points of aggregation to increase the amount of path sharing and reduce energy consumption. Data centric reinforcement mechanism is used to construct an energy efficient tree. Inefficient paths are excluded using greedy approach. Latency and energy savings are optimized but data accuracy is not considered. Wei et al. [38] proposed prediction based data aggregation technique by using double-queue mechanism to enable synchronization between sensor node and sink node to reduce cumulative error of continuous predictions. Three algorithms are designed: grey-model-based data aggregation (GMDA) is used to process data series noise, Kalman-filter based data aggregation (KFDA) is used to predict accuracy and combined grey model and Kalman filter data aggregation (CoGKDA) is used to increase prediction accuracy. The proposed technique reduces communication overhead and computational complexity but latency is not considered.

Yousefi et al. [98] proposed structure-free real-time data aggregation (RAG) protocol. Two mechanisms are considered for spatial and temporal convergence of packets. Judiciously waiting policy is used for temporal convergence and real-time data-aware anycasting policy is considered for spatial convergence of packets. The main objective is to improve energy consumption and end-to-end delay but network lifetime and communication overhead are not considered. Lin et al. [99] proposed evolutionary game-based data aggregation model (EGDAM). It is a data aggregation technique used to map the cooperation and competition in aggregation process into games through theoretic model to improve resilience. Based on this technique, evolutionary game based adaptive weighing algorithm (EGWDA) is proposed for pixel level data aggregation. The adjustment of weights between the sensors is considered as a game. An optimal weight distribution is achieved when a Nash equilibrium is achieved in weights allocation. In depletion zone, some nodes may not provide accurate information; therefore the proposed model takes the centroid value of dirty readings. Energy consumption and delay is larger in this scenario.

Paul and Gopinathana [86] proposed ACO based hybrid data aggregation mechanism which aggregates fruitful message through shortest path. Anomaly messages are identified and removed by a classification technique based on support vector machine. First of all, classification is done based on geographical location and the most powerful node is elected on the basis of computational capacity. At the cluster level, ant colony is implemented in order to collect the data from other nodes. Then messages are classified as anomaly or fruitful messages. Fruitful messages are forwarded further to base station for future decisions. This technique improves network life time but classification and communication cost is larger.

Lu et al. [85] proposed semi-structured multi-objective tree based dynamic data aggregation protocol in which ACO is used to find the optimal route for data transfer. Further, concept of sliding window is used for prediction of arriving packets to improve the aggregation probability and to reduce the transmission delay. For data aggregation, this protocol also shows that convergence of packet transmission from both temporal and spatial point of view. To achieve spatial convergence, DAA is proposed and RW is preferred for temporal convergence. Finally, simulation results validate the feasibility and the high efficiency of the novel protocol when compared with other existing approaches. Villas et al. [37] proposed scalability based data aggregation mechanism i.e. dynamic and scalable tree (DST) to create a scalable and dynamic structure to reduce the problem of load balancing. This technique decrease the number of messages required to establish a routing tree and chooses route with maximum rate of aggregation. It also maximizes the number of overlapping routes without considering energy cost. In this approach, there is one *Collaborator*, whose main task is to detect an event. *Coordinator*s gather data from collaborators and aggregate the data. A *relay node* is a node which forwards the data towards base station. *Aggregator nodes* forward the aggregated data from two or more coordinators. Latency can be included as future parameter.

Tree based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 17. *Energy* based autonomic data aggregation uses local minimum spanning tree (LMST) and relative neighborhood graph (RNG) topologies to construct a minimum spanning tree to calculate the position of next adjacent hop and to find the shortest path between sender and receiver to increase the network lifetime. *Adaptive energy aware* data aggregation uses sleep and awake technology in which parent and communicating nodes are in awake state while other nodes remain in sleep state. *Energy oriented cluster* based data aggregation divides sensor nodes of a cluster into two categories: simple relay nodes and network coder nodes (level of data correlation is used to aggregates data). *Collective tree* protocol based data aggregation collects and aggregates

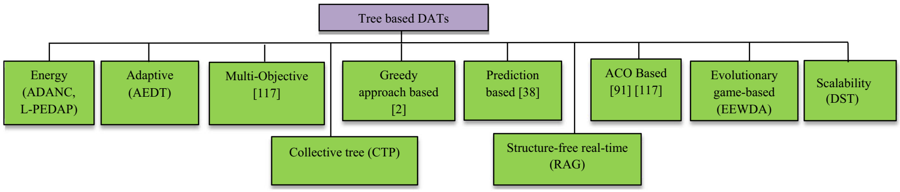


Fig. 17 Tree based DATs taxonomy

data using spanning tree which is an efficient method for data aggregation using beacon packets. *Greedy approach* based data aggregation is used to adjust points of aggregation to increase the amount of path sharing, energy consumption and improved latency and energy saving. *Prediction* based data aggregation uses double-queue mechanism to enable synchronization between sensor node and sink node to reduce cumulative error of continuous predictions. *Structure-free real-time* data aggregation uses *real-time data-aware anycasting and judiciously waiting policy* for spatial and temporal convergence of packets to improve energy consumption and end-to-end delay. *Evolutionary game*-based pixel-level data aggregation is used to map the cooperation and competition in aggregation process into games through theoretic model to improve resilience. Ant colony optimization based *hybrid* data aggregation aggregates fruitful message through shortest path and anomaly message has been removed by a classification technique based on support vector machine. Multiple objective trees based *dynamic* data aggregation uses ACO to find the optimal route for data transfer. In *scalability* based data aggregation, dynamic and scalable tree is used to reduce the problem of load balancing.

5.2.11 Network Density Based Data Aggregation Techniques

Data aggregation research work based on network density has been done by following authors. Intanagonwiwat et al. [28] proposed tree based data aggregation technique in which greedy approach is used to adjust points of aggregation to increase the amount of path sharing and reduce energy consumption. Data centric reinforcement mechanism is used to construct an energy efficient tree. Inefficient paths are excluded using greedy approach. Latency and energy saving are optimized but data accuracy is not considered.

Chen et al. [55] proposed cluster based adaptive data aggregation (ADA) technique, in which the main focus is to shift the load of sensor nodes and cluster heads to the resource enriched sink side and to enhance reliability factor of network. Sink node determines the reporting frequency of nodes and aggregation frequency of cluster heads. The reliability of event features is represented by two reliability parameters namely *spatial reliability* and *temporal reliability*. Spatial reliability is determined by aggregation ratio whereas temporal reliability is determined by reporting frequency. It achieves the desired reliability, but a lot of effort is required in terms of two dimensional parameters of spatial–temporal correlation.

Wang et al. [100] proposed a data aggregation technique to extend the network lifetime of WSNs using mobile relays by studying the performance of large dense network. Heterogeneous networks are considered as collection of both mobile and static nodes. It has been proved that lifetime of network using a mobile node is four times better than network using static nodes. Mobile nodes release the load of highly loaded network within a two- hop radius of base station and enhance the network lifetime. A joint mobility and routing algorithm has been proposed. Only some limited number of nodes is having the location of mobile node and hence decrease the computational cost.

Network Density based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 18. *Tree* based data aggregation uses greedy approach to adjust points of aggregation to increase the amount of path sharing and reduce the energy consumption. *Cluster based adaptive* data aggregation is performed at sink node in which aggregation ratio at CHs controls degree of spatial aggregation and reporting frequency at sensor nodes controls degree of temporal aggregation. *Network lifetime* based data aggregation is used to extend the network lifetime of WSNs using mobile relays by analyzing the performance of large dense network.

5.2.12 Prediction Based Data Aggregation Techniques

Data aggregation research work based on prediction has been done by following authors. Wei et al. [38] proposed prediction based data aggregation technique by using double-queue mechanism to enable synchronization between sensor node and sink node to reduce cumulative error of continuous predictions. Three algorithms are designed: grey-model-based data aggregation (GMDA) is used to process data series noise, Kalman-filter based data aggregation (KFDA) is used to predict accuracy and combined grey model and Kalman filter data aggregation (CoGKDA) is used to increase prediction accuracy, reduce communication overhead and computational complexity but latency is not considered. Jiang et al. [101] proposed energy-efficient data aggregation technique in which adaptive mechanism is used to enable or disable the prediction of performance. Further, analysis of performance is done to reduce prediction cost and communication cost and improve cluster-to-cluster propagation. For rapid propagation of aggregates, cluster to cluster propagation is used instead of node to node propagation.

Meng et al. [102] presented clustering based data aggregation technique DACP in which entire network is divided by sink node into different clusters and cluster head node is selected from every cluster to predict the data coming from different sources. In Prediction phase, the received predicted data is compared with sensed data to decide that whether this data should be send further or not. This technique improves network lifetime and reduces data transmission.

Prediction based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 19. *Double-queue mechanism* based data aggregation is used to enable synchronization between sensor nodes and sink to reduce cumulative error of continuous predictions. *Energy-efficient* data aggregation uses adaptive mechanism to enable or disable the prediction based on network condition. In *Clustering* based data aggregation, sink node divides the entire network into different clusters and cluster head node is selected from every cluster to predict the data coming from different sources and compares the predicted data with sensed data for further transmission of data.

5.2.13 Structure Free Data Aggregation Techniques

Structure Free data aggregation research work has been done by following authors. Yousefi et al. [98] proposed structure-free real-time data aggregation (RAG) protocol. Two mechanisms are considered for spatial and temporal convergence of packets. Judiciously Waiting policy is used for temporal convergence and real-time data-aware anycasting policy is considered for spatial convergence of packets. The main objective is to improve energy consumption and end-to-end delay but network lifetime and communication overhead are not considered.

Dietzel et al. [103] proposed fuzzy logic based approach for structure-free data aggregation technique for those applications which require periodic dissemination of

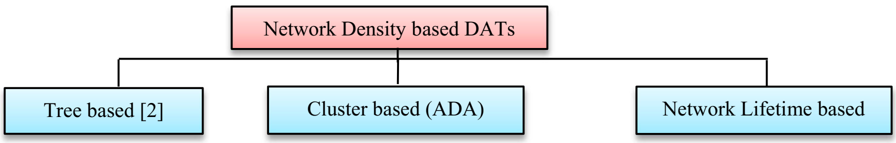


Fig. 18 Network density based DATs taxonomy

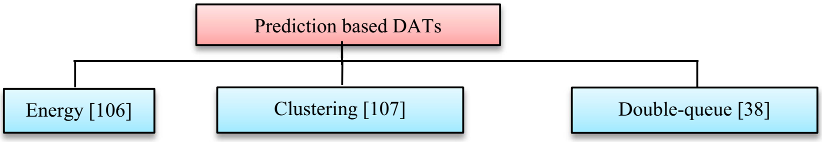


Fig. 19 Prediction based DATs TAXONOMY

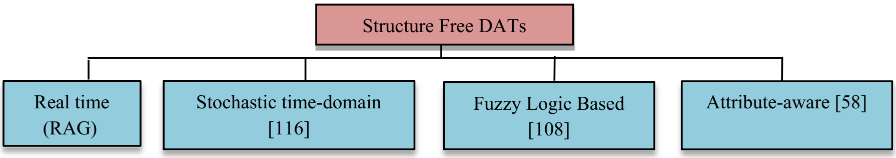


Fig. 20 Structure free DATs taxonomy

information to take efficient decisions. Fuzzy logic technique is used to take different decisions based on flexible and extensible set of inputs and outputs to satisfy the different requirements of applications. The decision is made on the basis of correlation in the data. Haghighi et al. [104] presented a stochastic time-domain method for burst aggregation of data in IEEE 802.15.4 protocol under the assumption of arrival of burst traffic which helps to find a station level information (failure distributions, collision and delay) using bottom-up approach. Two metrics i.e. transmission success rate and average packet loss have been designed through which performance of proposed method has been tested. Window sizes greater than 1 are also supported. The proposed approach is optimized in terms of computational complexity and memory.

Ren et al. [45] proposed packet-driven timing algorithm based attribute aware data aggregation (ADA) technique. In this, data can be gathered from heterogeneous nodes. As the data is gathered from different heterogeneous source nodes, attributes of packets are also different. So, an attribute aware data routing is recommended in this literature which makes the packets with the same attribute spatially convergent. To support ADA strategy, a hybrid approach of pheromone in ant colony and potential in physics is used. This technique improves scalability, node density, aggregation ratio and also considers mobility factors.

Structure Free based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 20. *Structure-free real-time* data aggregation uses *real-time data-aware anycasting* and *judiciously waiting policy* for spatial and temporal convergence of packets to improve energy consumption and end-to-end delay. *Fuzzy logic* based data aggregation is designed for those applications which require periodic dissemination of information to take efficient decisions. *Stochastic time-domain* method is designed for burst aggregation of data in IEEE 802.15.4 protocol which helps to find a station level information (failure distributions, collision and delay) using bottom-up approach. *Attribute aware* data aggregation uses packet-driven timing algorithm to aggregate data gathered from heterogeneous nodes.

5.2.14 Evolutionary Game Based Data Aggregation Techniques

Data aggregation research work based on evolutionary game has been done by following authors. Engouang et al. [70] presented game based data aggregation mechanism by considering ZigBee and Self-Fulfilling Belief (SFB) for two player's game theory. Tree and star based topologies are used to improve energy consumption. The main objective is to provide collision free communication to the nodes. Claude Shannon's information theory is considered for maintaining privacy. Based on SFB, every player plays its best with its energy parameter and also assumes that other players are also trying their best to win the game. This technique improves trustworthiness, power consumption, end-to-end delay and throughput but network lifetime and communication overhead are not considered.

Lin et al. [99] proposed evolutionary game-based data aggregation model (EGDAM). It is a data aggregation technique used to map the cooperation and competition in aggregation process into games through theoretic model to improve resilience. Based on this technique, evolutionary game based adaptive weighing algorithm (EGWDA) is proposed for pixel level data aggregation. The adjustment of weights between the sensors is considered as a game. An optimal weight distribution is achieved when a Nash equilibrium is achieved in weights allocation. In depletion zone, some nodes may not provide accurate information; therefore the proposed model takes the centroid value of dirty readings. Energy consumption and delay is larger in this scenario.

Evolutionary Game based Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 21. *Game* based data aggregation considers ZigBee and using self-fulfilling belief (SFB) to provide security and two player's game theory. Evolutionary game-based *pixel-level* data aggregation is designed to map the cooperation and competition in aggregation process into games through theoretic model to improve resilience.

5.2.15 Hybrid Data Aggregation Techniques

Hybrid data aggregation research work has been done by following authors. Paul and Gopinathana [86] proposed ACO based hybrid data aggregation mechanism which aggregates fruitful message through shortest path. Anomaly messages are identified and removed by a classification technique based on Support Vector Machine. First of all, classification is done based on geographical location and the most powerful node is elected on the basis of computational capacity. At the cluster level, ant colony is implemented in order to collect the data from other nodes. Then messages are classified as anomaly or fruitful messages. Fruitful messages are forwarded further to base station for future decisions. This technique improves network life time but classification and communication cost is larger.

Jung et al. [105] proposed hybrid clustering based data aggregation technique in which based on network status, adequate clustering technique is selected to improve data aggregation. In the proposed approach, BS broadcasts a control message to the network. Each node that receives this control message assumes sender node as parent node. There is one hop_count field in the control message which is used by the receiving node to calculate

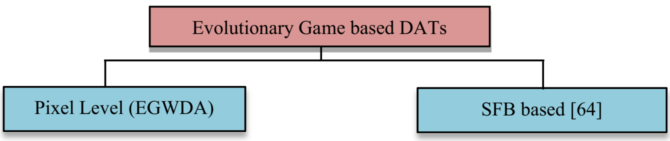


Fig. 21 Evolutionary game based DATs taxonomy

its distance from the BS. This message is again broadcasted until all the nodes in the network join the tree. A cluster head is elected using static random head selection technique and a timer is set for random time interval. When this timer expires, a node declares itself a cluster head and broadcasts this notification in the network and a cluster is formed in this way. This technique improves data transmission ratio and energy consumption by using static clustering in this approach.

Kim et al. [106] presented energy-aware hybrid data aggregation mechanism (EHDAM) in which data transmission is controlled adaptively with transmission based on timeout and burst length. The main problem addressed in this proposed work is *Energy Hole* problem. This problem arises due to the heavy burden near the sink and cluster heads, due to which node failure rate is higher in these regions. This technique is capable to adjust burst length threshold value to improve network lifetime and reduce the number of data transmission.

Hybrid Data Aggregation Taxonomy Based on above literature, following taxonomy has been derived as shown in Fig. 22. *Ant colony optimization* based hybrid data aggregation aggregates fruitful message through shortest path and anomaly message has been removed by a classification technique based on Support Vector Machine. *Clustering* based data aggregation is a clustering technique that can be used for target tracking applications based on network status. *Energy-aware* data aggregation controls data transmission adaptively based on timeout and burst length.

5.3 Comparison of Data Aggregation Techniques

Comparison of data aggregation techniques is a difficult task due to different types of data aggregation mechanisms and the lack of benchmarks. Therefore comparison of data aggregation techniques is significant to find the effective data aggregation technique. Table 4 shows the comparison of data aggregation techniques. Different traits of data aggregation techniques are considered and discussed below.

5.3.1 Traits of Data Aggregation Techniques

Data aggregation techniques in WSN can be compared based on some common characteristics for solving aggregation problems. Data aggregation technique, subtype, routing protocol, objective function, optimize parameter, merits, environment and demerits/open issues are some of the common and basic characteristics which should be examined in each data aggregation technique as described in Table 3.

5.4 Network Based Data Aggregation Techniques

In WSN, the process of data aggregation is done in two different types of networks namely Hierarchical and Network Flow based as shown in Fig. 23.

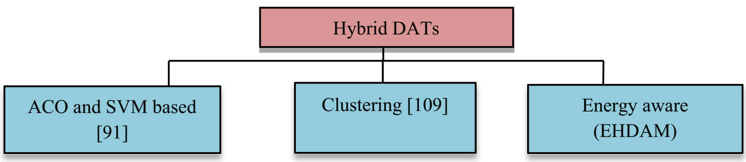


Fig. 22 Hybrid DATs taxonomy

Table 4 Comparison of data aggregation (DA) techniques

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
Adaptive DA	Energy and latency	Multihop	To reduce latency and energy consumption	Energy cost, Latency and delivery ratio	Power consumption reduced and reliability are increased	NS2	Network lifetime is lesser
	Application	Multihop	To decrease power and delay	End to end delay, energy and degree of aggregation	Overall reduction in header overhead, improves channel scheduling	Berkeley's MICA motes platform	Data accuracy is lesser
	Cluster	Hierarchical	To improve fault tolerance	Reliability	Reliability is increased	MATLAB	Problem of spatial–temporal correlation
	Energy	Tree	To enhance network lifetime and minimize energy consumption	Network lifetime, delay and power	Achieves good delivery ratio with reduced delay	NS2	Latency is not reduced
	SCT	Tree	To reduce the congestion and also the communication cost	Communication cost and data accuracy	Accuracy and efficiency are improved	GloMoSim	Network lifetime is lesser
	Distributed and energy	Cluster	To reduce energy consumption	Network lifetime	Improved energy consumption and lifetime	PROWLER	Data delivery latency is not considered
	Robust	Cluster	To improve robustness	Overall cost, duty cycle and delay	Energy efficiency is improved	TinyOS	Network lifetime is lesser
Cluster based DA	Adaptive	Hierarchical	To improve fault tolerance	Reliability	Reliability increased	MATLAB	Problem of spatial–temporal correlation

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Distributed	Chain	To reduce communication cost	Compression ratio	Communication cost is reduced	NS2	Energy consumption is not considered
	Hybrid	Multihop	To improve energy efficiency	Energy consumption, network lifetime, aggregation ratio and packet transmission success ratio	Enable Dynamic aggregation and improve network lifetime and energy consumption	Qual-Net	Latency is larger
	Head selection scheme	Chain	To reduce energy consumption	Network lifetime	Energy efficiency and network lifetime are improved	Omnet++ 4.0	Latency and communication cost are not considered
	Two tier clustering	Multihop	To minimize computation and communication Cost	Energy consumption, packet rate and throughput	Packet count and redundancy reduced, power consumption are improved	Network based simulator	Data accuracy is not considered and desired throughput are not achieved
	Balanced clustering	Multihop	To improve network lifetime and energy consumption	Network lifetime and power consumption	Energy efficiency and network lifetime are improved	Network based simulator	Latency is larger
	Group of clusters	Cluster	To improve network stability	Energy cost	Energy consumption reduced	Network based simulator	Data accuracy is not considered
	Energy	Chain	To improve network lifetime	Convergence rate, aggregation cycles, average packet drops, transmission cost and network lifetime	Transmission cost and network lifetime is improved	NS2	Lack of energy efficiency

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Adaptive and energy	Cluster	To reduce energy consumption	Network lifetime	Improved energy consumption and lifetime	PROWLER	Data delivery latency is not considered
	Bandwidth	Multihop	To improve energy efficiency	Throughput and power consumption	Packet delivery rate is reduced	Network based simulator	Desired energy efficiency is not achieved
	Load distribution	Distributed	To improve load distribution	Network lifetime	Network Lifetime and load distribution are improved	MATLAB	Power consumption is not considered
	Bandwidth and energy	Distributed	To improve throughput and packet delivery ratio	Energy consumption	Energy consumption, throughput and packet delivery ratio are improved	NS-2	Network Lifetime is not considered
	Diffusion and clustering	Cluster	To improve latency and reduce memory usage	Latency and memory usage	Redundancy is reduced	Network based simulator	Power consumption is not considered
	Hierarchical	Tree	To improve energy consumption	Energy cost	Energy consumption is reduced	SIDnet-SWANS	Distributed compressed sensing introduces complexity
Concealed DA	Hierarchical	Multihop	To improve trustworthiness	Data confidentiality and integrity	Data accuracy and communication cost improved	TinyOS 2.0	Latency is not considered
	Recoverable	Distributed	To reduce communication cost	Transmission overhead	Performance is improved	MICAz	Not consider transfer and data cost
	Multiple-application	Cluster	To improve robustness	Reliability and energy cost	Energy efficiency is improved	TinyPEDS	Data accuracy and latency is not considered

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Trust management	Multihop	To increase the level of security	Running time	Improved accuracy of data aggregation	Tiny OS 2.0	Desired power saving is not achieved
	Balance privacy-preserving	Distributed	To enhance the privacy-preserving efficacy	Communication overhead and privacy-preserving efficacy	Privacy-preserving efficacy is improved	Network based simulator	Network lifetime is lesser
	Privacy awareness	Multihop	To reduce the communication load	Congestion control and energy consumption	Improved security	Castalia simulator	Desired power saving is not achieved
Energy based DA	Secure pattern	Cluster	To improve the security	Occupied bandwidth rate and energy consumption	Trustworthiness is enabled	GloMoSim	Network lifetime is not considered
	Hierarchical	CDPR	To reduce energy consumption	Power and network lifetime	Network lifetime is improved	Testbed based simulator	Data accuracy and latency are not considered
	Co-operative	Cluster	To minimize power usage	Network lifetime	Energy consumption is improved	Testbed based simulator	Network lifetime is lesser
	Distributed and latency	Tree	To improve latency	Latency and energy consumption	Transmission cost and network lifetime are improved	Testbed based simulator	Desired power saving is not achieved
	Compressed	Chain	To reduce energy cost	Energy cost	Energy efficiency is improved	CPLEX	Network lifetime is lesser
	Data accuracy	Chain	To improve network lifetime	Network lifetime	Improved security, energy efficiency and data accuracy	NS2	Energy consumption is larger

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Tree based	Cluster	To reduce energy consumption	Energy cost	Energy efficiency is improved	Network based simulator	Network lifetime and communication overhead are not considered
	Privacy based	Tree	To reduce energy cost	Aggregation accuracy, data integrity, computation overhead, communication overhead	Energy consumption is improved slightly	Network based simulator	Energy consumption and communication cost are larger
	Design of structure based	Chain	To improve network lifetime and latency	Aggregation ratio, maintenance overhead and energy	Energy efficiency and network lifetime are improved	NS2	Data accuracy is not considered
	Game based	Tree	To improve trustworthiness and reduce power consumption	Control traffic received, Data Dropped, End-to-end delay and throughput	End-to-end delay and throughput are improved	OPNET	Network lifetime and communication overhead are not considered
	Reliability based	Cluster	To improve data accuracy	Network lifetime, energy and data accuracy	Improved security and energy efficiency	TOSSIM & TinyOS	Latency is not considered
	Static clustering	Cluster	To reduce energy cost	Energy consumption and data accuracy	Energy consumption is improved slightly	NS2	Energy consumption and delay are larger

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Delivery delay based	Cluster	To improve latency and network lifetime	Network lifetime and latency	Energy efficiency and network lifetime are improved	Discrete event-based simulator	Data accuracy is not considered
	Multi-objective meta-heuristic	Cluster	To improve security and energy efficiency	Threshold-based degree of intrusions	Energy efficiency and degree of intrusions are improved	Network based simulator	Network lifetime is not considered
	Remote component binding	Multihop	To reduce resource consumption and energy consumption	Energy consumption and latency	Energy consumption and latency are improved	OMNET++	Network lifetime is not considered
	Delay-aware	Tree	To reduce aggregation cost	Aggregation delay and energy efficiency	Aggregation cost and energy efficiency are improved	C++ based network simulator	Network lifetime is lesser
	Heterogeneous network	Multihop	To improve energy consumption and data quality	Precision and energy consumption	Energy consumption is reduced	MATLAB	More latency
	Dynamic	Chain	To improve energy cost	Scalability and energy	Energy cost is reduced	TOSSIM & TinyOS	Data accuracy and Latency is not considered
Latency based DA	Adaptive and energy	Multihop	To reduce latency and energy consumption	Energy Cost, Latency and Delivery ratio	Power consumption reduced and reliability are increased	NS2	Network lifetime is lesser
	Distributed	Cluster	To minimize latency	Aggregation latency	Transmission cost and network lifetime are improved	Network based simulator	Energy consumption and delay is larger
	Conflict aware	Tree	To minimize latency	Schedule latency	Latency is improved	Network based simulator	Delivery delay is larger

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
Network Lifetime based DA	Energy	Chain	To improve energy cost	Energy and network lifetime	Energy cost is reduced	Testbed based simulator	Latency is larger
	Linear programming	Cluster	To reduce the running time	Energy cost	Energy cost is reduced	Network based simulator	Energy consumption and communication cost are larger
	Precision constrained	Multihop	To improve network lifetime	Energy and network lifetime	Energy efficiency and network lifetime are improved	Ns2	Data accuracy and latency is not considered
	Distributed	Tree	To reduce the data traffic	Aggregated data rate and network lifetime	Data traffic is reduced	Network based simulator	Network lifetime and communication overhead are not considered
	Shortest path based	Tree	To improve network lifetime	Latency and energy consumption	Improved energy cost	Network based simulator	Energy consumption and delay is larger
	Multi-criterion decision-making	Cluster	To improve network lifetime and energy efficiency	Energy and time	Network lifetime and energy efficiency are improved	TinyOS	Energy efficiency is lesser
	Learning automata based	Multihop	To improve power consumption and network lifetime	Network time and energy	Time and energy consumption are improved	JSIM simulator	Load balancing is not effective

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	End-to-end delay	Cluster	To improve network lifetime	Average energy consumed per data packet and packet loss probability	Node Density, Average energy consumed per data packet and packet loss probability, End-to-end delay and network lifetime are improved	OPNET simulator	Impact of varying network load on the network lifetime is not analyzed
Network density based DA	Tree	Tree	To increase the amount of path sharing	Energy consumption	Latency is maintained and energy is saved	Network based simulator	Data accuracy is not considered
	Cluster	Hierarchical	To improve fault tolerance	Reliability	Reliability is increased	MATLAB	Problem of spatial–temporal correlation
Nature-inspired optimized DA	Search space based	Tree	To improve energy consumption	Cost	Transmission cost and network lifetime are improved	MATLAB	Communication overhead is not considered
	Energy	Cluster	To Reduce power consumption	Energy cost	Energy consumption is improved	Network based simulator	Data accuracy is not considered
	Network lifetime	Cluster	To extend the network lifetime	Energy efficiency, computation complexity and success Ratio	Energy cost is decreased	JProwler	Network lifetime and communication overhead are not considered
	Ladder diffusion	Distributed	To avoid wasted power and large processing time	Power consumption	Energy efficiency and network lifetime are improved	Network based simulator	Data accuracy is not considered

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Tree	Tree	To find the optimal route for data transfer	Aggregation probability and transmission delay	Aggregation probability and transmission delay are improved	C++	Energy consumption is not considered
	Hybrid	Tree	To reduce the congestion and also the communication cost	Communication cost, data accuracy	Accuracy and efficiency Improvement	GloMoSim	Network lifetime is lesser
QoS based DA	Energy	Cluster	To improve delay and data accuracy	Energy, accuracy and delay	Reduced delay	Network based simulator	Latency is not considered
	Waterfalls random partial aggregation	Cluster	To reduce delay and latency	Data Accuracy, Energy and Delay	Waiting time improved	Network based simulator	Energy consumption is larger
	Network lifetime and fault tolerance	Tree	To maximize network lifetime	Network lifetime	Transmission cost and network lifetime are improved	J-Sim	Energy consumption and delay is larger
	Time and energy	Chain	To ensure reliability and congestion avoidance	Throughput, latency, average energy consumption and average network lifetime	Throughput, latency improved	Network based simulator	Energy consumption and communication cost are larger
	Cost	Cluster	To find a low cost path for data transfer	Entropy and aggregation cost	Energy Cost is reduced	Network based simulator	Network lifetime is lesser
Scheduling based DA	Dynamic	Distributed	To minimize latency	Energy	Energy efficiency improved	OPNET	Larger network lifetime

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Distributed	Cluster	To improve network lifetime	Time latency and running time	Running time reduced	C++	More energy cost
	Semi-structured and unstructured	Cluster	To reduce energy consumption	latency and network lifetime	Energy cost reduced	NS2	Energy consumption and delay is larger
	Improved distributed	Distributed	To generate a collision-free schedule	Latency	Energy efficiency and network lifetime are improved	–	Data accuracy is not considered
	Collision free		To improve Fault tolerance	Latency and network size	Improved reliability	JProwler	Lesser data accuracy
	Delay based	Multihop	To improve the delay performance	End to End delay	Transmission cost and network lifetime are improved	NS2	Data accuracy is not considered
	Integrated tree construction	Tree	To find collision-free schedule	Execution time and Time latency	Improved communication overhead	TinyOS-2	Network lifetime is lesser
	Timeout control scheme	Tree	To reduce delivery delay	Energy and delay	Waiting time is improved	TinyOS	Not achieved desired energy consumption
Tree based DA	Energy	Cluster	To increase the network lifetime	Network lifetime and energy consumption	Energy consumption is reduced	Network based simulator	More latency
	Adaptive and energy	Tree	To enhance network lifetime and minimize energy consumption	Network lifetime and delay, power	Achieves good delivery ratio with reduced delay	NS2	Latency is not reduced
	Hybrid	Tree	To reduce the congestion and also the communication cost	Communication cost, data accuracy	Accuracy and efficiency Improvement	GloMoSim	Network lifetime is lesser

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
	Collective tree protocol	Cluster	To reduce communication overhead	Number of messages transmitted	Energy consumption is reduced	NS2	Network lifetime is lesser
	Energy oriented cluster based	Cluster	To improve energy consumption and network lifetime	Energy cost	Energy cost is reduced	TinyOS-2	Data delivery latency is not considered
	Greedy approach	Tree	To increase the amount of path sharing	Energy consumption, latency and energy saving	energy consumption, latency and energy saving are improved	NS2	Data accuracy is not considered
	Prediction based	Tree	To enable synchronization between sensor node and sink node	Prediction accuracy, communication overhead and computational complexity	Cumulative error of continuous predictions is reduced	TinyOS-2	Latency is not considered
	Structure-free real-time	Cluster	To improve energy consumption, and end-to-end delay	Aggregation gain, miss ratio, energy consumption, and end-to-end delay	Overall performance is improved	NS2	Network lifetime and communication overhead are not considered
	Evolutionary game-based	Cluster	To improve resilience	Latency and aggregation rate	Reasonable weights distribution of sensors can be achieved	C++	Energy consumption and delay is larger
	Dynamic	Tree	To find the optimal route for data transfer	Aggregation probability and transmission delay	Aggregation probability and transmission delay are improved	C++	Energy consumption is not considered
	Scalability	Tree	To maximize the number of overlapping routes	Communication costs and aggregation rate	Selects routes with the highest aggregation rate	Network based simulator	Not considered energy cost

Table 4 continued

Data aggregation (DA) technique	Subtype	Routing protocol	Objective function	Optimizing parameter	Merits	Operational environment	Demerits/open issues
Prediction based DA	–	Data collection	To increase prediction accuracy, reduce communication overhead, and computational complexity	Communication overhead and energy cost	Network lifetime and communication redundancy are improved	TinyOS-2	Latency and data accuracy is not considered
Structure-free DA	Real time	Cluster	To improve energy consumption, and end-to-end delay	Aggregation gain, miss ratio, energy consumption, and end-to-end delay	Overall performance is improved	NS2	Network lifetime and communication overhead are not considered
	Stochastic time-domain	Multihop	To improve network lifetime	Transmission success rate and average packet loss	Network lifetime is improved	C++	Energy consumption is not considered
	Attribute-aware	Chain	To improve scalability	Node density and aggregation ratio	Improved aggregation ratio	TOSSIM & TinyOS	Power consumption is not considered
Evolutionary game-based DA	–	Cluster	To improve resilience	Latency and aggregation rate	Reasonable weights distribution of sensors can be achieved	C++	Energy consumption and delay is larger
Hybrid DA	–	Tree	To improve network life time	Network lifetime	Reducing the congestion Problem	Weka	Energy consumption and communication cost are larger

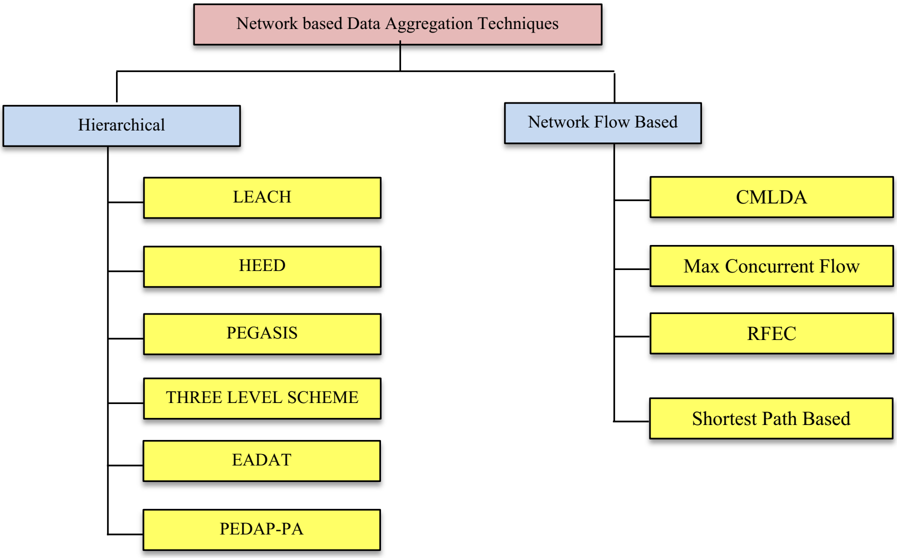


Fig. 23 Network based data aggregation techniques

In hierarchical network, aggregation of data has been done at special node, which is used to reduce energy consumption and number of data packets to destination [6, 11, 107, 108]. There are three types of networks being used in this network. First one is *cluster based*, in which data is transmitted from one cluster to another cluster through cluster heads after data aggregation, example is LEACH. Second is *chain based* in which, data is transmitted to closer neighbor node after the structuring of nodes into linear chain using greedy algorithm for data aggregation (For Example: PEGASIS). and third one is *tree based*, in which all the nodes are structured in the form of tree and transfer the data from leaf node to root node through intermediate node for data aggregation (For Example: EADAT) [7]. The classification of hierarchical data aggregation techniques [7, 8, 91] is summarized in Table 5.

In network flow based data aggregation, network architecture is designed to solve the network flow problem modeled as data aggregation. The objective of this technique is to improve network lifetime and energy consumption along with reducing flow constraints on information [13]. The classification of network flow based data aggregation techniques [109–111] is summarized in Table 6. Data aggregation techniques have been classified based on routing protocol and QoS parameters like latency, energy efficiency and data accuracy [2, 6, 14–16, 18, 20, 26, 27, 32, 66, 76, 97, 107, 108, 112, 113] and described in Table 7.

6 QoS-Aware Autonomic Cloud Based Wireless Sensor Network for Data Aggregation: A Solution

In WSNs, there is too much complexity for the development of applications (which depends on network protocols) due to limited memory, computational power and battery power of sensors. Data gathered should be managed in an efficient manner so that it can be

Table 5 Hierarchical data aggregation techniques

Technique	Objectives	Organization type	Features
LEACH [7]	Network lifetime is number of alive nodes	Cluster	Randomized rotation of cluster head (CH) Non-uniform energy consumption in various sensors
HEED [121]	Network lifetime is number of rounds till the death of first node	Cluster	Many energy levels in sensors Performs better than LEACH CHs are well distributed
PEGASIS [122]	Network lifetime is average of energy consumed by a node	Chain	Requirement of overall knowledge of network Less energy consumption as compared to LEACH
CHAIN BASED [122]	Network lifetime is multiplication of power consumption and delivery delay	Chain	Three level scheme better than PEGASIS and Binary chain based scheme better than LEACH in terms of performance
EADAT [123]	Network lifetime is number of live sensor nodes till complete transmission of data	Tree	Sink initiated broadcasting technique For broadcasting help messages select the threshold power
PEDAP-PA [7]	Network lifetime is till the expire of last node	Tree	Minimum spanning tree based technique Performs better than LEACH, PEGASIS

Table 6 Network flow based data aggregation techniques

Technique	Objectives	Methodology	Open issues
CMLDA [13]	Improve network lifetime, capacity and energy consumption	Integer linear programming technique	For networks of large size it results in high computational complexity
Max concurrent flow [124]	Improve energy consumption only	Dijkstra's shortest path tree technique	Rise in the number of sinks decreases performance gain
RFEC [125]	Improve number of rounds of data gathering to capacity and edge constraints	An approach to define the data flow based on residual graph	There is need of realistic models such as dynamic environments where sensors generate number of packets in a round based on requirement
Shortest path based [126]	Reduce transmission cost	Transmission structure based on Slepian–Wolf coding policies	How to establish the validity of Gaussian random field model?

used in future for different types of forecasting. For that, gathered data should be classified, stored and managed in an effective manner so that it can be easily accessed in future. But there is not enough space available to store gathered data for long time to make future decisions. The prediction of resources is required to manage gathered data in order to save cost and time to process large amount of data. To solve this problem, there is a need to use large cloud data centers which can provide large amount of space and improves reliability, efficiency and data retrieval rate etc.

Table 7 Tools and platforms used for data aggregation in WSNs

Tools	Description
NS2	NS-2 (Network Simulator Version2) is an open source simulation tool used for discrete events in WSN. NS-2 is implemented using C/C++ languages and OTCL. NS2 is easily maintainable and extensible. It is suitable for 802.14.4 and 802.11 type wireless MAC. NS-2 is an efficient tool for energy modeling in WSN
OMNeT++	OMNeT++ is a network simulator used for discrete events based on public source component and it is suitable for wireless and wired IP communication networks. Real time results can be represented in graphical form by using this tool
Castalia	This simulator is based on OMNeT++ platform and used for body area networks (BANs) and WSNs. It can also be used to test network of embedded devices with low power. Wireless channel model (for variation of path loss) and radio model (for low-power communication) can be used for testing distributed algorithms. Castalia provide provisions of sensing model to measure power consumption
Prowler	This simulator is used to design an event driven WSN to execute scenario in MATLAB. Prowler has ability to develop direct simulation code for routing protocol and can easily visualize the capabilities through user interface. This tool can also be used to test application codes used in wireless communication and simulate the propagation including collisions and radio transmission
J-prowler	J-prowler provides all the functionalities of Prowler as discussed but J-prowler is implemented in Java. Communication protocols of TinyOS can be analyzed and verified
TinyOS and TOSSIM	It is an open source simulator designed for wireless devices in sensor networks with low power consumption. TOSSIM is an emulator developed for wireless sensor network simulation on TinyOS for embedded operating system. Python is used to design TOSSIM and it can run easily on Linux Operating System. TOSSIM is used to simulate radio models and code executions but only homogeneous applications can be simulated
GlomoSim	This tool uses parsec compiler for parallel discrete event simulation. GlomoSim is able to simulate heterogeneous wireless network consist of more than 1000 nodes and it has ability to simulate multi-hop wireless communications using ad-hoc networking
Qual-Net	It is an extended version of GlomoSim, mostly used for defense projects and used for distributed applications and heterogeneous networks. Qual-Net provide graphical user interface to represent the real time results and speed of simulator can be increased by adding more processors
MATLAB	MATLAB (Simulink) is used to simulate various modulation & encoding techniques, physical layer parameters and communication channel in WSNs. It has capability to measure effect of different Noise schemes; SNR etc. in WSN and simulated results can be recorded and monitored easily using C and Java programming language. MATLAB (Simulink) has ability to extend the WSN by adding or removing end nodes and sensors
TinyPEDS	Tiny Persistent Encrypted Data Storage (tinyPEDS) is used in WSN to provide encryption during data aggregation. Through symmetric privacy holomorphic encryption, the stored data is secured after aggregation
UWSim	UWSim is a simulator used for Underwater Sensor Networks (UWSN). UWSim provides ability to handle scenarios specific to UWSN environments, for example limited bandwidth, less memory, high transmission energy and low frequency. It is implemented using network component-based approach, rather than a 3 layer/protocol-based approach

Table 7 continued

Tools	Description
Castalia	This simulator is based on OMNeT++ platform and used for Body Area Networks (BANs) and WSNs. It can also be used to test network of embedded devices with low power. The distributed algorithms are tested in wireless channel model (to variation of path loss) and radio model (for low-power communication) in Castalia. Castalia provides extensions for sensing modeling to measure power consumption. Implementation of sensor nodes in Castalia is done as compound modules which are used to represent different sub-modules for sensor, network based applications and network stack layers. Castalia is built using C++ language that provides features like physical process modeling and can used to implement MAC and routing protocols
EmStar	Linux-based framework, EmStar provides interface between deployment for iPAQ-class sensor nodes and simulation. There are two types of simulation environments which can be designed in EmStar namely *Real deployment* and *pure simulation*. Simulations for Mica2 motes and iPAQ based microservers can be developed using EmStar. EmStar adopts half simulation and half-emulation approach in which host machine is used to run software and interfacing with the real sensor. EmStar enables platform independent code and can execute on different platforms. For providing scalability, components can be dynamically added or removed in EmStar. Parallel simulations cannot be executed in EmStar
VisualSense	VisualSense uses component-based approach to construct the different wireless network to analyse energy consumption in WSN, media access control protocols, ad-hoc networking protocols and communication channels. Based on energy propagation model, VisualSense provides an effective radio model which can be reused for physical phenomena. VisualSense uses propagation model to provide sound model, which can be effectively used for localization
COOJA (COntiki Os JAva)	For Contiki sensor node OS (Operating system), COOJA simulator is used to provide functionality of high-level behavior in a single simulation platform and low level simulation of sensor node hardware. To support adaptability, CooJA allows different parameters of system can be changed like platforms of radio models, radio transceivers, OS and sensor node. Java Native Interface (JNI) is used in COOJA to make interaction with compiled code of Contiki. Contikicode can be executed by COOJA in two ways: compiling Contiki program for the MSP430 hardware or compiling the code of Contiki program directly on the host CPU. COOJA requires number of calculations to provide final result and it is difficult to use due to unavailability of GUI interface

The objective behind the development of this solution is to design a QoS aware cloud based data aggregation (CBDA) framework which can be used to gather, analyze and store the data of different domains coming from various sensors used in WSNs. The main objectives of CBDA framework are: to use sensors to collect data, to perform clustering of sensed data, to apply classification to categorize the data in different classes for easy interpretation, to identify different QoS requirements to process data and to use resources efficiently to perform data aggregation and store data on cloud data center for future use [21]. Zig Bee Protocol is used to tap the sensed and gathered data by various sensors and to further report that data to routers for data aggregation. Data can be in any form, like it may represent the conditions of environment like pressure, sound and temperature etc.

The architecture of QoS aware cloud based data aggregation framework is shown in Fig. 24. CBDA framework works in following steps: (1) use sensors to collect data, (2) clustering of sensed data is performed, (3) apply classification to categorize the data in different classes for easy interpretation and (4) identify different QoS requirements used to process data and use resources efficiently to perform data aggregation and store data on

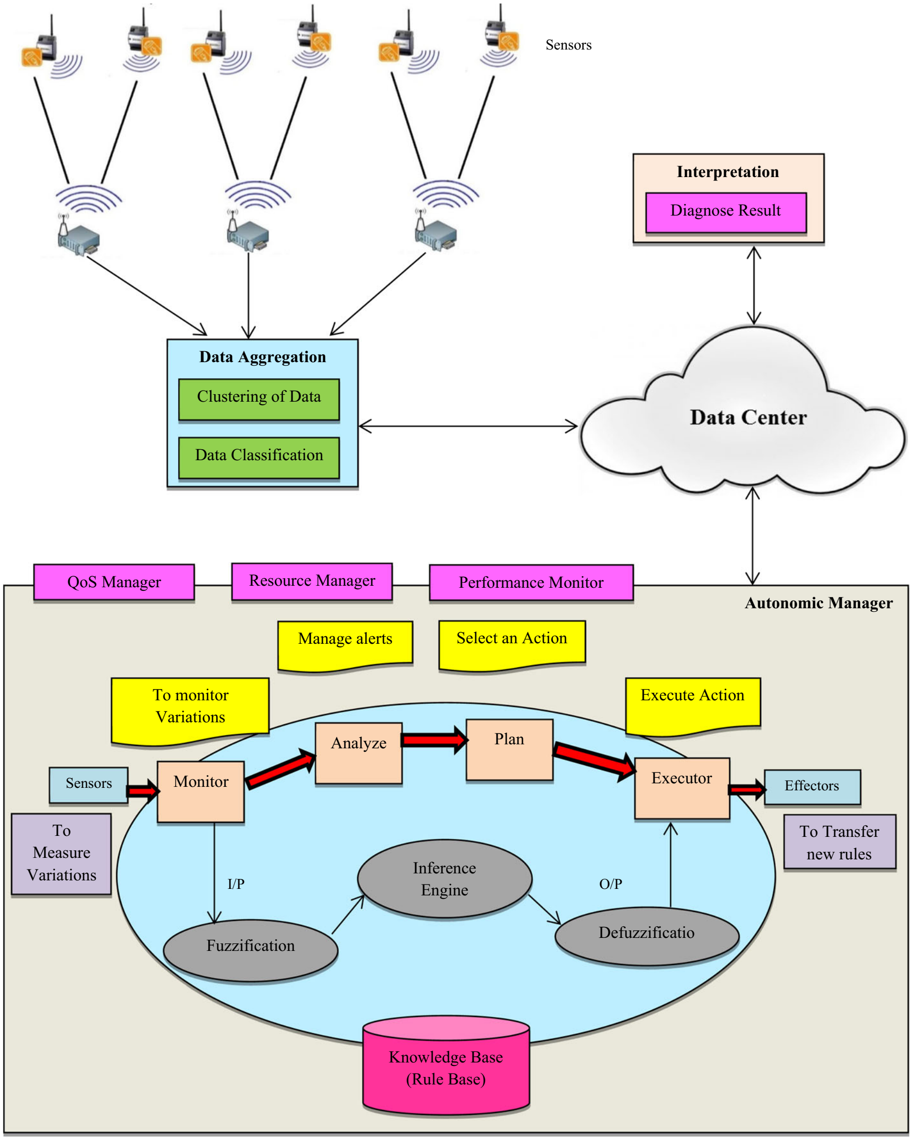


Fig. 24 QoS aware cloud based data aggregation (CBDA) framework

Cloud datacenter for future use [21]. Firstly unsupervised machine learning technique, i.e. *k*-means based clustering algorithm is used to *cluster* the unknown data coming from various sensors in WSN. K-means based clustering algorithm is a non-hierarchical method that initially takes the set of data of the population (Data Set (D)) and map that data into various clusters based on the value of centroids. The required number of clusters is decided in such a way that the centroids of selected points should be as far as possible.

Next, it examines the Data Set (D) in the population and assigns it to one of the clusters depending on its minimum distance from the centroid of the cluster. The cluster centroid's

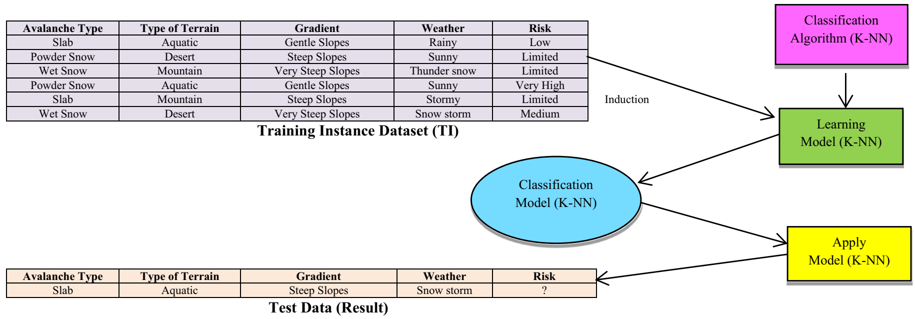
(C) position is recalculated every time a data is added to the cluster and this continues until all data points in a data set (D) are grouped and converged into required number of clusters (C(n)). *k*-means based clustering algorithm calculates the distance between a data reading and select that pair which shows the minimum distance and that particular reading and removes it from the actual data set (D) and assign it to some cluster. This whole process is repeated for the whole data set (D) until it converges to desired number of clusters (m). By performing the clustering process, final clusters of data (CD) have been created such as CD = $\{CD_1 \ldots CD_m\}$, *where* $m \ll t$. After this process, large number of data clusters is created which are difficult to understand and interpret. To solve this problem, classification technique is used to categorize the data in different classes for ease of interpretation.

Classification of the clustered data is required for further processing of data. CBDA framework gathers various types of data of different domains and classified them based on the domains specified in the database. The subtasks of information gathering and processing in CBDA framework are: (1) preprocessing, (2) attribute extraction and (3) classification. The number of sensors is sending their data of different domains from which system selects only relevant information and maintains this as a gathered target data. In this sub process, target datasets are created based on the relevant information that will be considered for further analysis in next subtasks. Elimination of irrelevant information will reduce the processing time.

Different sensors have different information regarding different domains. To develop a final training set, there is need of preprocessing steps because data might have missing data or noise issues. For critical evaluation, required number of samples have collected and analyzed. *Data cleaning* is performed to remove the inconsistent data, noisy data and to fill the data in missing values because dirty data will create confusion. In case of missing values, data is calculated by using weights, in which weight is assigned to particular values in fixed time intervals and missing values are filled by using adjacent values of that particular value. For noisy data (some error or variation in data), clustering technique is used which categorize the similar values in different clusters. Data constraints have been used to check the consistency of data and data is corrected manually to remove the inconsistency [21].

In CBDA framework, non-uniform data are converted into uniform data through data interpolation techniques. *Data integration* is used to combine the data coming from different sensors into a single data store. Through data integration different data is integrated but it contains some redundant data also. After this, data transformation is performed to convert the integrated data into an adequate format which is suitable for data mining. Normalization and aggregation techniques are used for data transformation in this research work. Further, processing of data will consume large time due to some complexity and redundancy in data. To eliminate these problems, *data reduction* is also performed to produce the quality of knowledge without compromising the integrity of the original data to identify effective analytical results. In this process, redundant data, irrelevant data or timely relevant data (which is not further required) are detected and removed.

Data transformation provides an interface between data analysis sub process (classification) and data preprocessing [21]. After data preprocessing this process converts the labeled data into an adequate format which is suitable for classification. In *data preprocessing*, data may be presented in different formats. It is very common that real life data consider more variables than required to classify the information. The main aim of this sub process is to reduce the effective number of variables. In CBDA framework, lossless aggregation technique is used to present data in recognizable format after merging and data reduction. CBDA framework considers different types of variables and classifies these



Avalanche Type	Type of Terrain	Gradient	Weather	Risk
Slab	Aquatic	Gentle Slopes	Rainy	Low
Powder Snow	Desert	Steep Slopes	Sunny	Limited
Wet Snow	Mountain	Very Steep Slopes	Thunder snow	Limited
Powder Snow	Aquatic	Gentle Slopes	Sunny	Very High
Slab	Mountain	Steep Slopes	Stormy	Limited
Wet Snow	Desert	Very Steep Slopes	Snow storm	Medium

Training Instance Dataset (TI)

Avalanche Type	Type of Terrain	Gradient	Weather	Risk
Slab	Aquatic	Steep Slopes	Snow storm	?

Test Data (Result)

Fig. 25 Classification process

variables based on their domains and store the corresponding information in the cloud data center. Based on this classification, CBDA framework can automatically diagnose the aggregated data. *Principal component analysis* (*PCA*) is used to find the distinct attributes to reduce the correlation among attributes. After analysis, the stored data are transformed into a new storage location and now data is very easily distributed into different classes. Mathematically, variance by any projection of data is calculated in PCA and decides its coordinates [21]. The first coordinate (first principal component) contains the projection of data with greater variance, second coordinate (second principal component) contains the projection of data with second greatest variance and so on. Principal component (PC) is the smallest number of uncorrelated variables derived from correlated variables through transformation in the PCA.

After attribute extraction of data, it is required to classify the gathered information of different users of different domains [21]. K-NN (k-nearest neighbor) classification mechanism is used to identify the different class labels of users. K-NN is supervised machine learning technique which is used to classify the unknown data using a training data set generated by it. The training data set includes defined class labels and their similar properties. In K-NN algorithm, the distance is computed from one specific instance to every training instance to classify that unknown instance. Both *k*-nearest neighbor and *k* minimum distance is determined and output class label is identified among *k* classes. During the training phase, K-NN algorithm utilizes training data. After classification, the required data can be traced easily from database. Figure 26 illustrates the diagnose process after classification used in this research work.

As shown in Fig. 25, K-NN model is used to identify the class of risk through training instance dataset (TI).Test data acts as an input of this model and it is compared with TI and identifies the class in which data is laid using following rule:

$$\text{Rule: } \mathbf{If}\ \{Avalanche\,Type\ \wedge\ Type\ of\ Terrain\ \wedge\ Gradient\ \wedge\ Weather\}\ \mathbf{then}\ Risk$$

CBDA framework processes and aggregates the data. To perform data aggregation and manage the large data in cloud data center, there is need of efficient management of resources. Firstly, the *QoS manager* predicts the QoS requirements to perform the entire process. Based on QoS information, resource requirement is predicted. After resource requirements prediction, resources are allocated to perform data aggregation. After allocation of resources, actual processing of data is started [21]. During execution of large data, performance is monitored continuously to maintain the efficiency of CBDA framework and generates alert in case of performance degradation. Alerts can be generated in three conditions generally: (1) if resource consumption is more than threshold values of resource consumption, ii) if energy consumption is more than threshold values of energy consumption and (2) if the number of required resources is more than the number of the provided resources. Working of autonomic manager of CBDA is based on IBM's autonomic model that considers four steps of autonomic system: monitor, analyze, plan and execute as shown in Fig. 24.

Sensors get the information about performance and current status of other nodes being used in the system. Firstly, the updated information from processing nodes is transferred to manager node and then manager node transfers this information to other sensors. Updated information includes information about QoS parameters (execution time, execution cost, resource consumption and energy consumption). Initially, *monitor* collects the information from sensors for analyzing performance variations by comparing expected and actual performance. Actual information about performance is observed based on QoS parameter readings and transfers this information to the next module for further analysis. For monitoring resource requirements, energy consumption and resource consumption, QoS agent is installed on all processing nodes to monitor the performance. A set of resources ($A_R = \{R_1, R_2, \ldots R_o\}$) is considered and placed in resource list for data processing. If ([Required Resources (RR) < Provided Resources (PR)] || [Resource Consumption (RC_q) > Threshold Value (TV_R] && [Energy Consumption ($Energy_{consumption}$) > Threshold Value ((TV_E)]) then alert will be generated otherwise resources are scheduled for data processing.

Analyze and plan module start analyzing the information received from monitoring module and make a plan for adequate actions for corresponding alert. Energy consumption, resource consumption and resource requirement are calculated and resources for execution are allocated. CBDA compares the value of energy consumption and resource consumption with a threshold value. If resource consumption is less than resource threshold value and energy consumption is less than the energy threshold value and if the number of required resources is less than provided resources, then aggregation of data continues otherwise no resource is allocated and process of reallocation is started. After meeting this condition, resources are allocated for further processing of data and the value of energy consumption, resource consumption and resource requirement are checked periodically as shown in Fig. 26. In case, sensed value is greater than the threshold, an alert will be generated by the system automatically. Same corrective measures are performed twice and if CBDA fails to correct it still, then the system will be treated as down.

CBDA framework always executes the resources with highest reliability. The components of fuzzy system used in this framework are described as.

It is necessary to find the input and output parameters for fuzzy systems. Basic information includes execution time, cost, resource name and resource type, but these parameters are constant, does not affect the final output. Fuzzy inputs include energy consumption ($Energy_{consumption}$) and resource consumption (RC_q) and fuzzy output is reliability (R). All the three parameters are categorized into three levels: low, medium and high. All the three variables changes continuously according to the conditions of the region of interest and hence these variables are considered [21]. Further, all the input and output variables are classified into various groups. Every group represents the fuzzy set of given input or output. Based on this, fuzzy rule set is created to define the behavior of fuzzy system and relationship between inputs and outputs. Maximum energy consumption and resource consumption are fixed in this framework; otherwise it will generate alert.

In this framework, three *membership functions* are considered for every input and output variables: Low, Medium and High. Based on these input and output variable, the inference

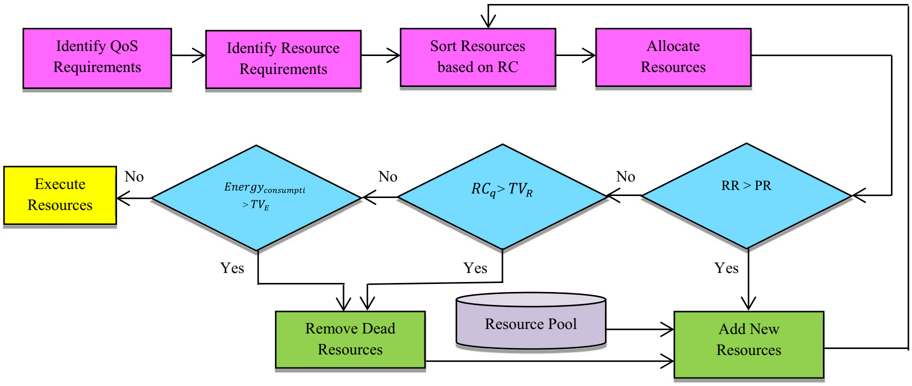


Fig. 26 Autonomic execution of resources in CBDA

engine is making decisions. The value of membership functions can be changed based on the requirements and conditions of every workload. In the fuzzy logic system, the working of inference engine is similar to the reasoning process of human. Energy consumption ($Energy_{consumption}$) and resource consumption (RC_q) are antecedents and reliability (R) is consequent in this framework. Three membership functions are considered for every two inputs. Several rules are being used simultaneously in the inference process. *Fuzzification* is used to find the degree of truth for every rule; membership function defined for every input variable is applied to their actual value. *Defuzzification* is used to convert the value of fuzzy output into crisp output value. MAXIMUM method is used in this work for Defuzzification. The crisp output is selected for output variable for which the value is maximum.

Executor implements the plan after analyzing it completely and it is used to analyze the performance of management of resources. The main task of the executor is to reduce the energy consumption and resource consumption. Based on the output given by the analysis, executor tracks the request submission and takes the action according to the rules described in the knowledge base. Effectors are used to exchange updated information and to transfer the new policies, rules and alerts to other nodes with updated information. Executor will add new node from resource pool with minimum resource consumption and energy consumption. If the resources are not available in resource pool then add new node from reserve resource pool with minimum resource consumption and energy consumption. During processing of data, performance is monitored continuously using a sub unit *performance monitor* to maintain the efficiency of system and generates alert in case of performance degradation. Alerts can be generated if the value of energy consumption and resource consumption is more than threshold value. If problem is not resolved, then alert will be generated and information will be sent to administrator.

7 Tools and Platforms used for Data Aggregation in WSNs

Table 7 shows the various platforms and tools used for data aggregation in WSNs. The most commonly used are NS2, MATLAB, OMNET++ and GlomoSim. Table 7 gives a description of the main features supported by the tools and platforms.

8 Discussions

123 research articles are reviewed in this research work and are presented in a systematic and categorized manner. This research article puts emphasis on data aggregation, evolution of data aggregation, types of data aggregation techniques and comparison of data aggregation techniques. Existing research by Rajagopalan and Varshney [11] and Jesus et al. [12] reflect research issues. Initial void for future work has been filled by these surveys in the domain of data aggregation. They identified only 11 main research studies in the area of data aggregation. Focus of our survey paper is wider than existing surveys and comprises the most recent research work related to data aggregation in WSN up to 2015 using the methodical survey technique [22]. The data aggregation techniques and their sub types in detail are explored. The data aggregation techniques are compared based on important aspects of data aggregation. The research issues are addressed and open challenges which are still unresolved in data aggregation techniques are explored. Furthermore, key discoveries in data aggregation techniques occurred after 2001 are found. In this research article, the existing research is presented in chronological order in different sections which provide easy understanding to the perspective readers and authors to find the latest research done after 2007. This methodical analysis has suggestions for perspective research scholars who are already working in this area and looking for new ideas and professional experts working in data aggregation. A lot of open issues are presented for professional experts and perspective researchers. The design of data aggregation techniques for a specific network will be depend on the aims and objectives.

9 Extraction Outcomes

The motive of this research work is to find the available research in data aggregation algorithms in WSNs. Figure 27 shows the percentage of research paper discussing different data aggregation algorithms (adaptive, cluster, concealed, energy, latency, network lifetime, network density, nature-inspired optimized, QoS, scheduling, tree, prediction, structure free real time, evolutionary game and hybrid) from year 2002 to 2015 [114–116]. Maximum research work has been done in the area of energy based data aggregation while data aggregation with prediction, structure free real time, evolutionary game and hybrid data aggregation techniques are still an emerging area. Most explored area of research in WSN is cluster based data aggregation after energy based data aggregation.

Figure 28 describes the percentage of research papers which are considering different QoS parameters (network life time, reliability, congestion control, energy, data accuracy and latency) from year 2002 to 2015.

Maximum research work has been done in data aggregation in WSN by considering energy as a QoS parameter while reliability, congestion control considered as a QoS parameter still requires lot of efforts. Figure 29 shows the number of research articles published from year 2002 to 2015 in the area of data aggregation in WSN. Maximum research has been done on data aggregation in year 2011 and minimum in year 2002 and 2005.

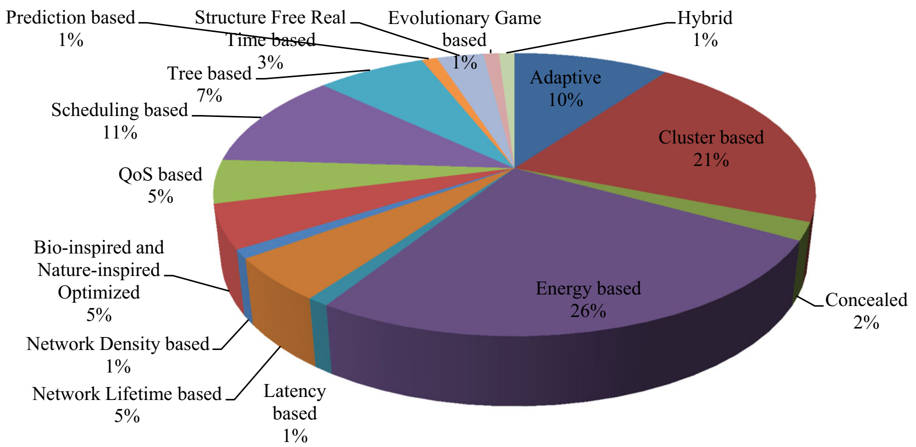
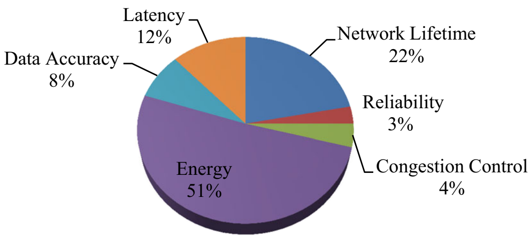


Fig. 27 Data aggregation algorithms in WSNs

Fig. 28 QoS parameters considered in data aggregation Algorithms



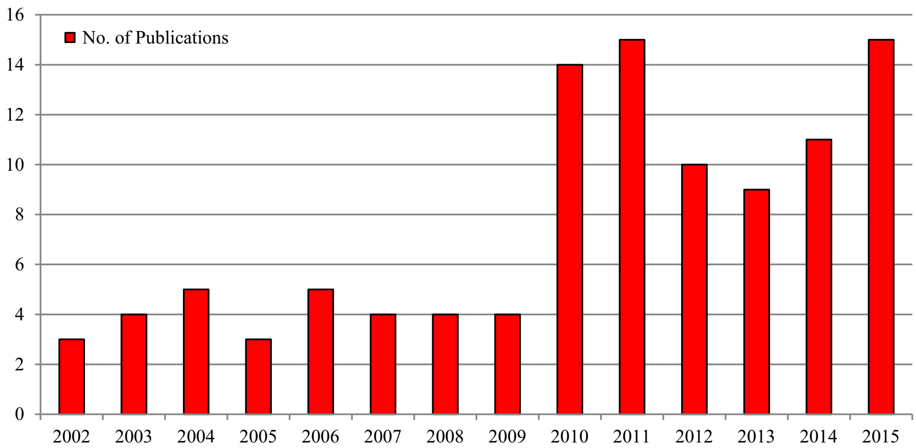


Fig. 29 Time count in data aggregation techniques

10 Conclusions: Future Research Directions

Research papers from existing research work have been studied. 123 out of 932 were identified to be the most suitable research papers of data aggregation techniques. In this research paper, the results have been analyzed in various ways like types of data aggregation techniques, data aggregation methodologies and tools and platforms used in WSNs for data aggregation. A comparison of data aggregation techniques has been presented. We have identified from existing literature that there is an excessive disparity in the definition of “data aggregation”. Contrast and assessment of types of data aggregation technique in WSNs can aid to select the adequate technique. Possible future directions can be:

(a) *Load Balancing and Data Aggregation* Under high network load conditions, traffic can be split over multiple paths which are discovered during route discovery phase using data dissemination algorithms. But, the concurrent use of multiple alternate paths results in inter-path interference which results in degradation of performance of the network. When multiple paths are concurrently used, then there must be some data aggregation or data fusion procedure, so that data can be merged together before sending to the host controller or sink. In this way resources like bandwidth, energy etc. can be utilized better.

(b) *Resource Consumption* In order to prolong the network lifetime in data aggregation, it is required that all the available resources are consumed efficiently in a network. Energy is the most crucial resource which must be consumed intelligently as this is the scarce resource. WSNs are usually deployed in harsh environments where power supply is not an option. Also, there is a tradeoff between energy efficiency and reliability. In multipath routing, multiple paths are discovered between the source and destination by exchanging the messages. Then data is transferred by using these multiple paths and the traffic is distributed in a uniform way. But maintaining multiple paths is itself a very critical task, especially in case of coding based multipath routing. So there must be a balance between the energy efficiency and reliability.

(c) *Quality of Service (QoS)* Quality of Service demands based on applications while performing data aggregation is an open research issue in WSNs. The various QoS parameters which can be considered are energy, delay, reliability etc. Some researchers made efforts to provide the required level of QoS parameters independent of the application.

(d) *Reliability and Fault-Tolerance* Due to time-varying network conditions and dynamic topology, providing reliability and fault tolerance is a very challenging task in data aggregation in WSN. This research issue can be analyzed in terms of lossy and lossless data aggregation.

(e) *Number of paths* The main issues in data aggregation are: how many paths are sufficient to provide reliability and fault tolerance; how paths are discovered and maintained and based on what parameters optimal paths are discovered. But there is a tradeoff between the required level of fault tolerance and route maintenance overhead.

10.1 Parameters to be considered for designing data aggregation technique

There are four aspects which must be considered while implementing data aggregation approaches: QoS Support, time bound, application oriented and energy efficient.

(a) *Time Bound* While performing data aggregation, time constraint should also be considered along with energy efficiency. During data aggregation, a considerable time is involved in aggregating the data and then reporting aggregated data to the base station. But, in some applications, delay is the main constraint. So an aggregation approach should be applied in such a manner that it reduces delay along with energy consumption.

(b) *Energy efficiency* Due to high node density in WSNs, generally same data is sensed by many sensors, which results in data redundancy due to spatial correlation. So, a lot of resources like energy, bandwidth, and time are wasted in case of redundancy. To alleviate this drawback, data fusion is used. In data aggregation process, data is send to elected aggregator nodes, which further aggregates the data and then send the same to the base station. It saves energy, as in traditional approach each node sends its readings to the base station individually.

(c) *Application Oriented* Data aggregation is applied by considering the application requirements. In some cases, delay is the main constraint, thus lossy aggregation may be applied in this case as data accuracy is not the main concern. In other applications, data accuracy is the main concern, so there is a necessity of lossless aggregation as we would like to grasp the minute details of every event.

(d) *QoS Support* Energy is one of the main constraints in WSNs. There are several other parameters also known as quality of service (QoS) parameters which are considered while designing and implementing data aggregation approach. Some QoS parameters are delay, data accuracy, bandwidth and throughput etc.

We hope that this research work will be beneficial for researchers who want to do research in any area concerning to WSNs such as data aggregation and impact of data aggregation on different QoS parameters [12, 117–120].

Appendix 1: Acronyms

WSN	Wireless Sensor Network
TAG	Tiny AGgregation
CH	Cluster Head
DAT	Data Aggregation Technique
FoS	Focus of Study
SCT	Semantic Correlation Tree
QoS	Quality of Service
LEACH	Low Energy Adaptive Clustering Hierarchy
PEGASIS	Power Efficient data GAthering protocol for Sensor Information Systems
EADAT	Energy Aware Distributed heuristic Aggregation Transmission
PEDAP-PA	Power Efficient Data gathering and Aggregation Protocol Power Aware
RFEC	Restricted Flow problem with Edge Capacities

DMAC	Delay aware Medium Access Control
HEED	Hybrid Energy Efficient Distributed Clustering Approach
AIDA	Application Independent Data Aggregation
CMLDA	Clustered Maximum Lifetime Data gathering with Aggregation
TDMA	Time Division Multiple Access
AEDT	Adaptive Energy aware Data aggregation Tree
TTCDA	Two Tier Cluster based Data Aggregation
GCEDA	Grouping nodes and Clusters for Efficient Data Aggregation
DASDR	Data Aggregation Supported by Dynamic Routing
DST	Dynamic and Scalable Tree
ESPDA	Energy efficient Secure Pattern based Data Aggregation
LMST	Local Minimum Spanning Tree
CTP	Collective Tree Protocol
KFDA	Kalman-Filter based Data Aggregation
GMDA	Grey-Model-based Data Aggregation
AEDT	Adaptive Energy aware Data aggregation Tree
DAACA	Data Aggregation Ant Colony Algorithms

References

1. Tan, H. Ö., & Korpeoglu, I. (2003). Power efficient data gathering and aggregation in wireless sensor networks. *ACM SIGMOD Record, 32*(4), 66–71.
2. Pourpeighambar, S. B., Aminian, M., & Sabaei, M. (2011). Energy efficient data aggregation of moving object in wireless sensor networks. In *Australasian telecommunication networks and applications conference* (pp. 1–8).
3. Krishnamachari, L., Estrin, D., & Wicker, S. (2002). The impact of data aggregation in wireless sensor networks. In *Proceedings of 22nd international conference of distributed computing system work* (pp. 575–578).
4. Qayyum, B., Saeed, M., & Roberts, J. A. (2015). Data aggregation in wireless sensor networks with minimum delay and minimum use of energy: A comparative study. In Accepted for publication in *Electronic Workshops in Computing* (*eWiC*). British Computer Society.
5. Cayirci, E. (2003). Data aggregation and dilution by modulus addressing in wireless sensor networks. *IEEE Communication Letters, 7*(8), 355–357.
6. Dagar, M., & Mahajan, S. (2013). Data aggregation in wireless sensor network: A survey. *International Journal of Information and Computation Technology, 3*(3), 167–174.
7. Tan, H. Ö., & Körpeoğlu, I. (2003). Power efficient data gathering and aggregation in wireless sensor networks. *ACM SIGMOD Record, 32*(4), 66–71.
8. Madden, S., Franklin, M. J. Hellerstein, J. M., & Hong, W. (2002). TAG: A tiny aggregation service for ad hoc sensor networks. In *Proceedings of 5th symposium operating systems design implementation* (Vol. 36, no. SI, pp. 131–146).
9. Al-Karaki, I. N., Ul-Mustafa, R., & Kamal, A. E. (2004). Data aggregation in wireless sensor networks—Exact and approximate algorithms. In *Work. High performance switching and routing*, 2004. HPSR (pp. 241–245).
10. Massad, Y. E., Goyeneche, M., Astrain, J. J. & Villadangos, J. (2008). Data aggregation in wireless sensor networks. In *3rd international conference information communication technologies from theory to applications* (Vol. 2, pp. 1040–1052).
11. Rajagopalan, R., & Varshney, P. K. (2006). Data-aggregation techniques in sensor networks: A survey. *IEEE Communications Surveys and Tutorials, 8*(4), 48–63.
12. Jesus, P., Baquero, C., & Almeida, P. S. (2015). A survey of distributed data aggregation algorithms. *IEEE Communications Surveys & Tutorials, 17*(1), 381–404.
13. Kalpakis, K., Dasgupta, K., & Namjoshi, P. (2003). Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. *Computer Networks, 42*(6), 697–716.

14. Lu, G., Krishnamachari, B., & Raghavendra, C. S. (2004). An adaptive energy-efficient and low-latency MAC for data gathering in wireless sensor networks. In *18th international parallel distributed processing symposium 2004 proceedings*, 2004.
15. Li, W., Bandai, M., & Watanabe, T. (2010). Tradeoffs among delay, energy and accuracy of partial data aggregation in wireless sensor networks. In *Proceedings of IEEE international conference advanced information networking and applications AINA* (pp. 917–924).
16. Li, H., Lin, K., & Li, K. (2011). Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks. *Computer Communications, 34*(4), 591–597.
17. Liu, C. X., Liu, Y., Zhang, Z. J., & Cheng, Z. Y. (2013). High energy-efficient and privacy-preserving secure data aggregation for wireless sensor networks. *International Journal of Communication Systems, 26*(3), 380–394.
18. Li, H., Wu, C., Hua, Q. S., & Lau, F. C. M. (2011). Latency-minimizing data aggregation in wireless sensor networks under physical interference model. *Ad Hoc Networks, 12,* 52–68.
19. Shan, M., Chen, G., Luo, D., Zhu, X., & Wu, X. (2014). Building maximum lifetime shortest path data aggregation trees in wireless sensor networks. *ACM Transactions on Sensor Networks, 11*(1), 11–18.
20. Tsai, S. Y., Sou, S. I., & Tsai, M. H. (2014). Reducing energy consumption by data aggregation in M2M networks. *Wireless Personal Communications, 74*(4), 1231–1244.
21. Randhawa, S., & Jain, S. (2017). An intelligent PSO-based energy efficient load balancing multipath technique in wireless sensor networks. *Turkish Journal of Electrical Engineering & Computer Sciences, 25*(4), 3113–3131.
22. Randhawa, S., & Jain, S. (2015). A systematic review on energy aware QoS routing in wireless sensor networks. *International Journal of Energy, Information and Communications, 6*(5), 1–14.
23. Al-Karaki, J. N., Ul-Mustafa, R., & Kamal, A. E. (2009). Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms. *Computer Networks, 53*(7), 945–960.
24. Li, M., Xu,, Wang, S., & Tang, S. (2009). Efficient data aggregation in multi-hop wireless sensor networks under physical interference model. In *IEEE 6th international conference on mobile adhoc and sensor systems* (pp. 353–362).
25. Rout, R. R., & Ghosh, S. K. (2014). Adaptive data aggregation and energy efficiency using network coding in a clustered wireless sensor network: An analytical approach. *Computer Communications, 40,* 65–75.
26. Mantri, D., Prasad, N. R., & Prasad, R. (2013). MHBCDA: Mobility and heterogeneity aware bandwidth efficient cluster based data aggregation for wireless sensor network. In *3rd International conference on wireless communications, vehicular technology, information theory and aerospace & electronics systems (VITAE)* (pp. 24–27).
27. Banerjee, R. (2014). Cluster based routing algorithm with evenly load distribution for large scale networks. In *2014 International conference on computer communication and informatics (ICCCI)* (no. I, pp. 1–6).
28. Intanagonwiwat, C., Estrin, D., Govindan, R., & Heidemann, J. (2002). Impact of network density on data aggregation in wireless sensor networks. In *Proceedings of 22nd international conference of distributed computing system* (pp. 17–18).
29. Chatterjea, S. (2003). A dynamic data aggregation scheme for wireless sensor networks. In *Proceedings of the 14th ProRISC workshop on circuits, systems and signal processing* (pp. 1–7). Japan: Kokurakita.
30. He, T., Blum, B. M., Stankovic, J. A., & Abdelzaher, T. (2004). AIDA: Adaptive application-independent data aggregation in wireless sensor networks. *ACM Transactions on Embedded Computing Systems (TECS), 3*(2), 426–457.
31. Hu, F., Cao, X., & May, C. (2005). Optimized scheduling for data aggregation in wireless sensor networks. *In International conference on information technology: Coding and computing*, *2005. ITCC 2005* (pp. 557–561).
32. Çam, H., Özdemir, S., Nair, P., Muthuavinashiappan, D., & Sanli, H. O. (2006). Energy-efficient secure pattern based data aggregation for wireless sensor networks. *Computer Communications, 29*(4), 446–455.
33. Gao, J., Guibas, L., Milosavljevic, N., & Hershberger, J. (2007). Sparse data aggregation in sensor networks. In *6th International conference on information processing in sensor* networks, *ACM Proceeding* (pp. 430–439).
34. Yu, B., Li, J., & Li, Y. (2009). Distributed data aggregation scheduling in wireless sensor network. In *IEEE INFOCOM 2009—28th conference of computation communication* (pp. 2159–2167).
35. Jiang, H., Jin, S., Wang, C., & Member, S. (2010). Parameter-based data aggregation for statistical information extraction in wireless sensor networks. *IEEE Transactions Vehicular Technology, 59*(8), 3992–4001.

36. Li, Y., Guo, L., & Prasad, S. K. (2010). An energy-efficient distributed algorithm for minimum-latency aggregation scheduling in wireless sensor networks. In *Proceeding of IEEE international conference distributed computing systems* (pp. 827–836).
37. Villas, L. A., Guidoni, D. L., Araujo, R. B., Boukerche, A., & Loureiro, A. F. (2010). A scalable and dynamic data aggregation aware routing protocol for wireless sensor networks. In *Proceedings of the 13th ACM international conference on modeling, analysis, and simulation of wireless and mobile systems* (pp. 110–117).
38. Wei, G., Ling, Y., Guo, B., Xiao, B., & Vasilakos, A. V. (2011). Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman Filter. *Computer Communications, 34*(6), 793–802.
39. Jung, W. S., Lim, K. W., Ko, Y. B., & Park, S. J. (2011). Efficient clustering-based data aggregation techniques for wireless sensor networks. *Wireless Networks, 17*(5), 1387–1400.
40. Guo, W., Xiong, N., Vasilakos, A. V., Chen, G., & Cheng, H. (2011). Multi-source temporal data aggregation in wireless sensor networks. *Wireless Personal Communications, 56*(3), 359–370.
41. Chen, C. M., Lin, Y. H., Lin, Y. C., & Sun, H. M. (2012). RCDA: Recoverable concealed data aggregation for data integrity in wireless sensor networks. *IEEE Transactions Parallel and Distributed Systems, 23*(4), 727–734.
42. Mantri, D., Prasad, N. R., Prasad, R., & Ohmori, S. (2012). Two tier cluster based data aggregation (TTCDA) in wireless sensor network. *IEEE International Conference Advanced Networks Telecommunciations Systems, 2012,* 117–122.
43. Kuo,T. W., & Tsai, M. J. (2012). On the construction of data aggregation tree with minimum energy cost in wireless sensor networks: NP-completeness and approximation algorithms. In *Proceedings of IEEE INFOCOM* (pp. 2591–2595).
44. Virmani, D., Sharma, T., & Sharma, R. (2013). Adaptive energy aware data aggregation tree for wireless sensor networks. *International Journal of Hybrid Information Technology, 6*, 26–36.
45. Ren, F., Zhang, J., Wu, Y., He, T., & Chen, C. (2013). Attribute-aware data aggregation using potential-based dynamic routing in wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems, 24,* 881–892.
46. Mantri, D., Prasad, N. R., & Prasad, R. (2013). Grouping of clusters for efficient data aggregation (GCEDA) in wireless sensor network. In *3rd IEEE International advance computing conference IACC 2013* (pp. 132–137).
47. Kumar, M., & Rajkumar, N. (2013). SCT based adaptive data aggregation for wireless sensor networks. *Wireless Personal Communications, 75*(4), 2121–2133.
48. Li, D., Zhu, Q., Du, H., & Li, J. (2012). An improved distributed data aggregation scheduling in wireless sensor networks. *Journal of Combinatorial Optimization, 27*(2), 221–240.
49. Mantri, D. S., Prasad, N. R., & Prasad, R. (2015). Bandwidth efficient cluster-based data aggregation for wireless sensor network. *Computers & Electrical Engineering, 41,* 256–264.
50. Lee, H., Hwang, H., Duc, T. L., Shon, M. H., Choo, H., & Kim, D. S. (2015). Restructuring binomial trees for delay-aware and energy-efficient data aggregation in wireless sensor networks. In *Proceedings of the 9th international conference on ubiquitous information management and communication* (pp. 13–20).
51. Liu, Y., Liu, C. X., & Zeng, Q. (2015). Improved trust management based on the strength of ties for secure data aggregation in wireless sensor networks. *Telecommunication Systems, 62*(2), 319–325.
52. Azad, P., & Sharma, V. (2015). Pareto-optimal clustering scheme using data aggregation for wireless sensor networks. *International Journal of Electronics, 102*(7), 1165–1176.
53. Asemani, M., & Esnaashari, M. (2015). Learning automata based energy efficient data aggregation in wireless sensor networks. *Wireless Networks, 21*(6), 2035–2053.
54. Kitchenham, B. A., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*. Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University, Keele and Department of Computer Science, University of Durham, Durham, UK (p. 65).
55. Chen, H., Mineno, H., & Mizuno, T. (2008). Adaptive data aggregation scheme in clustered wireless sensor networks. *Computer Communications, 31*(15), 3579–3585.
56. Wu, W., Cao, J., Wu, H., & Li, J. (2012). Robust and dynamic data aggregation in wireless sensor networks: A cross-layer approach. In *2012 9th International conference on ubiquitous intelligent computing* (Vol. 57, pp. 306–313).
57. Zheng, J., Member, S., Wang, P., & Li, C. (2010). Distributed data aggregation using Slepian–Wolf coding in cluster-based wireless sensor networks. *IEEE Transactions on Vehicular Technology, 59*(5), 2564–2574.

58. Maraiya, K., Kant, K., & Gupta, N. (2011). Efficient cluster head selection scheme for data aggregation in wireless sensor network. *International Journal Computer Applications, 23*(9), 10–18.
59. Yuea, J., Zhang, W., Xiao, W., Tang, D., & Tang, J. (2012). Energy efficient and balanced cluster-based data aggregation algorithm for wireless sensor networks. *Procedia Engineering, 29,* 2009–2015.
60. Sinha, A., & Lobiyal, D. K. (2013). Performance evaluation of data aggregation for cluster-based wireless sensor network. *Human-Centric Computing and Information Sciences, 3*(1), 1–13.
61. Xu, X., Ansari, R., Khokhar, A., & Vasilakos, A. V. (2015). Hierarchical data aggregation using compressive sensing (HDACS) in WSNs. *ACM Transactions on Sensor Networks (TOSN), 11*(3), 1–25.
62. Ozdemir, S., & Xiao, Y. (2011). Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. *Computer Networks, 55*(8), 1735–1746.
63. Lin, Y. H., Chang, S. Y., & Sun, H. M. (2013). CDAMA: Concealed data aggregation scheme for multiple applications in wireless sensor networks. *IEEE Transactions on Knowledge and Data Engineering, 25*(7), 1471–1483.
64. Zhang, C., Li, C., & Zhao, Y. (2015). A balance privacy-preserving data aggregation model in wireless sensor networks. *International Journal of Distributed Sensor Networks, 2015*, 1–10.
65. Sicari, S., Grieco, L. A., Boggia, G., & Porisini, A. C. (2012). DyDAP: A dynamic data aggregation scheme for privacy aware wireless sensor networks. *Journal of Systems and Software, 85*(1), 152–166.
66. Chen, Y. P., Liestman, A. L., & Liu, J. (2006). A hierarchical energy-efficient framework for data aggregation in wireless sensor networks. *IEEE Transactions Vehicular Technology, 55*(3), 789–796.
67. Xu, H., Huang, L., Zhang, Y., Huang, H., Jiang, S., & Liu, G. (2010). Energy-efficient cooperative data aggregation for wireless sensor networks. *Journal of Parallel and Distributed Computing, 70*(9), 953–961.
68. Xiang, L., Luo, J., & Vasilakos, A. (2011). Compressed data aggregation for energy efficient wireless sensor networks. In *8th Annual IEEE communications society conference sensor, mesh ad hoc communications networks* (pp. 46–54).
69. Chao, C. M., & Hsiao, T. Y. (2014). Design of structure-free and energy-balanced data aggregation in wireless sensor networks. *Journal of Network and Computer Applications, 37,* 229–239.
70. Engouang, T. D., Liu, Y., & Zhang, Z. (2014). GABs: A game-based secure and energy efficient data aggregation for wireless sensor networks. *International Journal of Distributed Sensor Networks, 501,* 1–31.
71. Liu, C., Liu, Y., & Zhang, Z. (2013). Improved reliable trust-based and energy-efficient data aggregation for wireless sensor networks. *International Journal of Distributed Sensor Networks, 2013,* 1–13.
72. Krishna, M. B., & Doja, M. N. (2015). Multi-objective meta-heuristic approach for energy-efficient secure data aggregation in wireless sensor networks. *Wireless Personal Communications, 81*(1), 1–16.
73. Ramachandran, G. S., Daniels, W., Proença, J., Michiels, S., Joosen, W., Hughes, D., & Porter, B. (2015). Hitch Hiker: A remote binding model with priority based data aggregation for wireless sensor networks. *In Proceedings of the 18th international ACM SIGSOFT symposium on component-based software engineering* (pp. 43–48).
74. Xiao, S., Li, B., & Yuan, X. (2015). Maximizing precision for energy-efficient data aggregation in wireless sensor networks with lossy links. *Ad Hoc Networks, 26,* 103–113.
75. Zhang, J., Wu, Q., Ren, F., He, T., & Lin, C. (2010). Effective data aggregation supported by dynamic routing in wireless sensor networks. *IEEE International Conference Communications, 2010,* 1–6.
76. Liu, H., Liu, Z., Li, D., Lu, X., & Du, H. (2013). Approximation algorithms for minimum latency data aggregation in wireless sensor networks with directional antenna. *Theoretical Computer Science, 497,* 139–153.
77. Xue, Y., Cui, Y., & Nahrstedt, K. (2005). Maximizing lifetime for data aggregation in wireless sensor networks. *Mobile Networks and Applications, 10*(6 SPEC. ISS), 853–864.
78. Tang, X., & Xu, J. (2006). Extending network lifetime for precision-constrained data aggregation in wireless sensor networks. In *Proceedings IEEE INFOCOM.*
79. Yum, S. P. (2008). Optimal routing and data aggregation for maximizing lifetime of wireless sensor networks. *IEEE/ACM Transactions on Networking, 16*(4), 892–903.
80. Awang, A., & Agarwal, S. (2015). Data aggregation using dynamic selection of aggregation points based on RSSI for wireless sensor networks. *Wireless Personal Communications, 80*(2), 611–633.
81. Misra, R., & Mandal, C. (2006). Ant-aggregation: Ant colony algorithm for optimal data aggregation in wireless sensor networks. In *IFIP international conference on wireless and optical communications networks* (pp. 1–5). Bangalore.

82. Yucheng, W. L., & Fan, K. C. (2007). An ant colony algorithm for data aggregation in wireless sensor networks. *In SensorComm international conference on sensor technologies and applications* (pp. 101–106).
83. Lin, C., Wu, G., Xia, F., Li, M., Yao, L., & Pei, Z. (2012). Energy efficient ant colony algorithms for data aggregation in wireless sensor networks. *Journal of Computer and System Sciences, 78*(6), 1686–1702.
84. Ho, J. H., Shih, H. C., Liao, B. Y., & Chu, S. C. (2012). A ladder diffusion algorithm using ant colony optimization for wireless sensor networks. *Information Sciences (NY), 192,* 204–212.
85. Lu, Y., Comsa, I. S., Kuonen, P., & Hirsbrunner, B. (2015). Dynamic data aggregation protocol based on multiple objective tree in wireless sensor networks. In *2015 IEEE tenth international conference on intelligent sensors, sensor networks and information processing* (*ISSNIP*) (pp. 1–7).
86. Paul, B., & Gopinathan, E. (2014). Hybrid data aggregation technique in wireless sensor network through classification of fruitful messages. In *Fourth international conference advances computing and communications* (pp. 157–175).
87. Pham, T., Kim, E. J., & Moh, M. (2004). On data aggregation quality and energy efficiency of wireless sensor network protocols—extended summary. In *Proceedings of first international conference broadband networks* (pp. 3–5).
88. Chen, I. R., Speer, A. P., & Eltoweissy, M. (2011). Adaptive fault-tolerant QoS control algorithms for maximizing system lifetime of query-based wireless sensor networks. *IEEE Transactions on Dependable and Secure Computing, 8*(2), 161–176.
89. Misra, S., & Thomasinous, P. D. (2010). A simple, least-time, and energy-efficient routing protocol with one-level data aggregation for wireless sensor networks. *Journal of Systems and Software, 83*(5), 852–860.
90. Chen, C., Lee, K., Park, J., & Baek, S. J. (2015). Minimum cost data aggregation for wireless sensor networks computing functions of sensed data. *Journal of Sensors*, 1–15.
91. Bagaa, M., Derhab, A., Lasla, N., Ouadjaout, A., & Badache, N. (2012). Semi-structured and unstructured data aggregation scheduling in wireless sensor networks. In *Proceedings of IEEE INFOCOM* (pp. 2671–2675).
92. Jhumka, A., Bradbury, M., & Saginbekov, S. (2014). Efficient fault-tolerant collision-free data aggregation scheduling for wireless sensor networks. *Journal of Parallel and Distributed Computing, 74*(1), 1789–1801.
93. Joo, C., Choi, J. G., & Shroff, N. B. (2010). Delay performance of scheduling with data aggregation in wireless sensor networks. In *IEEE proceedings INFOCOM*.
94. Bagaa, M., Younis, M., Djenouri, D., Derhab, A., & Badache, N. (2015). Distributed low-latency data aggregation scheduling in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN), 11*(3), 1–36.
95. Kwon, S., Ko, J. H., Kim, J., & Kim, C. (2011). Dynamic timeout for data aggregation in wireless sensor networks. *Computer Networks, 55*(3), 650–664.
96. Tan, H. O., Korpeoglu, I., & Stojmenovi, I. (2011). Computing localized power-efficient data aggregation trees for sensor networks. *IEEE Transactions on Parallel Distributed Systems, 22*(3), 489–500.
97. Hakoura, B., & Rabbat, M. G. (2012). Data aggregation in wireless sensor networks: A comparison of collection tree protocols and gossip algorithms. In *25th IEEE Canadian conference on electrical and computer engineering* (*CCECE*) (pp. 1–4).
98. Yousefi, H., Yeganeh, M. H., Alinaghipour, N., & Movaghar, A. (2012). Structure-free real-time data aggregation in wireless sensor networks. *Computer Communications, 35*(9), 1132–1140.
99. Lin, J., Xiong, N., Vasilakos, A. V., Chen, G., & Guo, W. (2011). Evolutionary game-based data aggregation model for wireless sensor networks. *IET Communications, 5*(12), 1691.
100. Wang, W., Srinivasan, V., & Chua, K. (2008). Extending the lifetime of wireless sensor networks through mobile relays. *IEEE/ACM Transaction Networking, 16*(5), 1108–1120.
101. Jiang, H., Jin, S., & Wang, C. (2011). Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor networks. *IEEE Transactions Parallel and Distributed Systems, 22*(6), 1064–1071.
102. Meng, L., Zhang, H., & Zou, Y. (2011). A data aggregation transfer protocol based on clustering and data prediction in wireless sensor networks. In *7th International conference wireless communications networking and mobile computing* (pp. 1–5).
103. Dietzel, S., Bako, B., Schoch, E., & Kargl, F. (2009). A fuzzy logic based approach for structure-free aggregation in vehicular ad-hoc networks. In *Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking VANET 09* (p. 79).

104. Haghighi, M. S., Xiang, Y., Varadharajan, V., & Quinn, B. (2015). A stochastic time-domain model for burst data aggregation in IEEE 802.15.4 wireless sensor networks. *IEEE Transactions on Computers, 64*(3), 627–639.
105. Jung, W. S., Lim, K. W., Ko, Y. B., & Park, S. J. (2009). A hybrid approach for clustering-based data aggregation in wireless sensor networks. In *2009 Third international conference on digital society* (pp. 112–117).
106. Kim, M. G., Han, Y. T., & Park, H. S. (2011). Energy-aware hybrid data aggregation mechanism considering the energy hole problem in asynchronous MAC-based WSNs. *IEEE Communications Letters, 15*(11), 1169–1171.
107. Chaudhury, B. P., & Nayak, A. K. (2015). Energy saving performance analysis of hierarchical data aggregation protocols used in wireless sensor network. In *Advances in intelligent systems and computing* (Vol. 309, pp. 79–89). Springer.
108. Saini, K., Kumar, P., & Sharma, J. (2013). A survey on data aggregation techniques for wireless sensor networks. *International Journal of Advanced Research in Computer Engineering & Technology, 3*(7), 901–903.
109. Xu, X., Li, X. Y., Mao, X., Tang, S., & Wang, S. (2011). A delay-efficient algorithm for data aggregation in multihop wireless sensor networks. *IEEE Transactions Parallel and Distributed Systems, 23*(1), 163–175.
110. Groat, M. M., Hey, W., & Forrest, S. (2011). KIPDA: k-indistinguishable privacy-preserving data aggregation in wireless sensor networks. In *Proceedings IEEE INFOCOM* (pp. 2024–2032).
111. Su, L., Gao, Y., Yang, Y., & Cao, G. (2011). Towards optimal rate allocation for data aggregation in wireless sensor networks. In *Proceedings of Twelfth ACM international symposium mobile ad hoc networking and computing—MobiHoc*.
112. Enachescu, M., Goel, A., Govindan, R., & Motwani, R. (2005). Scale-free aggregation in sensor networks. *Theoretical Computer Science, 344*(1), 15–29.
113. He, W., Nguyen, H., Liuy, X., Nahrstedt, K., & Abdelzaher, T. (2008). iPDA: An integrity-protecting private data aggregation scheme for wireless sensor networks. In *MILCOM 2008 IEEE military communications conference* (pp. 1–7).
114. Esnaashari, M., & Meybodi, M. R. (2010). Data aggregation in sensor networks using learning automata. *Wireless Networks, 16*(3), 687–699.
115. Huang, S. I., Shieh, S., & Tygar, J. D. (2010). Secure encrypted-data aggregation for wireless sensor networks. *Wireless Networks, 16*(4), 915–927.
116. Ozdemir, S., & Çam, H. (2010). Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks. *IEEE/ACM Transactions on Networking, 18*(3), 736–749.
117. He, W., Liu, X., Nguyen, H., Nahrstedt, K., & Abdelzaher, T. (2007). PDA: Privacy-preserving data aggregation in wireless sensor networks. In *IEEE INFOCOM 2007—26th IEEE international conference on computer communications* (pp. 2045–2053).
118. Patil, N. S., & Patil, P. R. (2010). Data aggregation in wireless sensor network. In *Proceedings of IEEE international conference computational intelligence and computing research* (pp. 28–29).
119. Tsitsipis, D., Dima, S. M., Kritikakou, A., Panagiotou, C., & Koubias, S. (2011). Data merge: A data aggregation technique for wireless sensor networks. In *IEEE 16th conference on emerging technologies & factory automation* (pp. 1–4).
120. Hamid, A., Ehsan, S., & Hamdaoui, B. (2014). Rate-constrained data aggregation in power-limited multi-sink wireless sensor networks. In *International wireless communications and mobile computing conference* (*IWCMC*) (pp. 500–504).
121. Lou, E., Hill, D. L., & Raso, J. V. (2010). HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *Medical Biological and Engineering Computing, 48*(3), 235–243.
122. Lindsey, S., Raghavendra, C., & Sivalingam, K. M. (2002). Data gathering algorithms in sensor networks using energy metrics. *IEEE Transaction on Parallel and Distributed Systems, 13*(9), 924–935.
123. Ding, M., Cheng, X., & Xue, G. (2003). Aggregation tree construction in sensor networks. In *2003 IEEE 58th vehicular technology conference VTC 2003-Fall (IEEE Cat. No.03CH37484*) (Vol. 4, pp. 2168–2172).
124. Xue, Y., Cui, Y., & Nahrstedt, K. (2005). Maximizing lifetime for data aggregation in wireless sensor networks. *Mobile Networks and Applications, Special Issue on Energy Constraints and Lifetime Performance in Wireless Sensor Networks, 10*(6), 853–864.
125. Hong, B., & Prasanna, V. K. (2004). Optimizing system life time for data gathering in network sensor systems. In *Proceeding of algorithms wireless and ad-hoc networks.*

126. Cristescu, R., Beferull-Lozano, B., & Vetterli, M. (2004). On network correlated data gathering. *IEEE INFOCOM, 4*(4), 2571–2582.

Sukhchandan Randhawa received M.E. degree in Computer Science and Engineering from Thapar University Patiala, India in 2012 She received B.E. degree in Computer Science and Engineering from Chitkara Institute of Engineering and Technology, Rajpura, Punjab, India in 2010. She had done Diploma in Information Technology from Thapar Polytechnic, Patiala, India in 2007. She is pursuing Ph.D. in Wireless Sensor Networks. Her research interests are focused on Wireless Sensor Networks, power efficiency, data aggregation and load balancing algorithms in Wireless Sensor Networks. She has joined Thapar University as Lecturer in 2012.

Sushma Jain received the B.E. degree in computer technology from S.A.T.I Vidisha India in 1993 and the Ph.D. degree in Computer Science, Thapar University from, Patiala, India in 2012 She worked as a Lecturer with the S.A.T.I., Vidisha from 1996 to 2001. She joined Thapar University, in Feb. 2001 and is currently working as an Assistant Professor at the Computer Science and Engineering Department in Patiala, India. Her research interests are focused on Wireless Sensor Networks, in particular development of data aggregation and load balancing algorithm and scheduling of resources.